GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: March 17, 2004, 21:14:07 ; Search time 1659.36 Seconds
(without alignments)
18464.115 Million cell updates/sec

Title: US-09-664-186-4
Perfect score: 1026
Sequence: 1 gtgaagaacgaaaaaacctt.....ccctcaaggaggcctcctga 1026

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	60	5.8	683	14	CA235176	CA235176 SCCCFL409
c	2	54.6	5.3	925	29	CNS0091P	AL053013 Drosophil
	3	53.2	5.2	655	29	CG352167	OG1BW12TV
c	4	53.2	5.2	679	29	CC609109	CC609109 OGWDJ44TH
c	5	53.2	5.2	773	29	CC661222	CC661222 OGLAU18TV
	6	53.2	5.2	807	28	BZ529007	BZ529007 OGAIR91TC
	7	53.2	5.2	815	29	CG034209	CG034209 PUJGY50TD
	8	53.2	5.2	897	29	CC661216	CC661216 OGLAU18TH
	9	51.8	5.0	925	29	CNS0091P	AL053013 Drosophil
	10	51.2	5.0	649	14	CA230834	CA230834 SCJFFL3C0
	11	51.2	5.0	649	14	CB870888	CB870888 HC15J04w
c	12	50.6	4.9	594	14	CB874255	CB874255 HC15J04y
	13	50.6	4.9	637	12	BI960118	BI960118 HVSMEn002
	14	50.6	4.9	664	13	BQ762267	BQ762267 EBro01_SQ
	15	50.6	4.9	973	14	CA275829	CA275829 SCCCSD109
	16	50.4	4.9	855	29	CG041883	CG041883 PUJEF24TD
	17	50.4	4.9	890	29	CC731258	CC731258 OGLAX50TH
	18	50.4	4.9	937	29	CC729649	CC729649 OGLAJ02TV
	19	50.2	4.9	509	12	BJ549094	BJ549094 BJ549094
	20	49.8	4.9	608	14	CA179624	CA179624 SCVPST106
	21	49.8	4.9	659	14	CA194749	CA194749 SCRFSB102
	22	49.8	4.9	694	14	CA196087	CA196087 SCSBAD108
	23	49.4	4.8	642	29	CG043714	CG043714 PUILF82TB
c	24	49.4	4.8	775	29	CG207191	CG207191 OGXFN36TH
	25	49.4	4.8	793	29	CG449321	CG449321 OGVHG34TH
c	26	49.4	4.8	837	29	CC701337	CC701337 OGUKL21TV
	27	49.4	4.8	1201	13	BX381961	BX381961 BX381961
c	28	49.2	4.8	502	12	BM335757	BM335757 MEST160-F
c	29	49.2	4.8	728	14	CF638161	CF638161 zmrww00_0
	30	49.2	4.8	759	10	BF259495	BF259495 HVSMEf001
	31	49	4.8	745	29	CC633883	CC633883 OGNAK47TV
c	32	49	4.8	747	29	CC633878	CC633878 OGNAK47TH
	33	49	4.8	869	29	CG284200	CG284200 OG0BZ21TV
c	34	49	4.8	909	29	CG228851	CG228851 OG2BX92TH
	35	48.6	4.7	559	14	CA208502	CA208502 SCMCSB111
	36	48.6	4.7	564	14	CA210811	CA210811 SCEPSB113
	37	48.6	4.7	605	14	CA181598	CA181598 SCBFST313
	38	48.6	4.7	620	14	CA182549	CA182549 SCCCST308
	39	48.6	4.7	647	14	CA216848	CA216848 SCRFAD112
	40	48.6	4.7	730	14	CA195888	CA195888 SCEZAD108
c	41	48.6	4.7	902	28	BZ579152	BZ579152 msh2_6221
	42	48	4.7	524	9	AV436476	AV436476 AV436476
	43	48	4.7	807	29	CC731270	CC731270 OGVAG66TV
c	44	48	4.7	910	29	CG323065	CG323065 OGWKZ61TV
	45	48	4.7	1066	14	CA254564	CA254564 SCBFFL411

ALIGNMENTS

RESULT 1
CA235176
LOCUS CA235176 683 bp mRNA linear EST 25-SEP-2003
DEFINITION SCCCFL4093A04.g FL4 Saccharum officinarum cDNA clone SCCCFL4093A04 5', mRNA sequence.
ACCESSION CA235176
VERSION CA235176.1 GI:35303456
KEYWORDS EST.
SOURCE Saccharum officinarum
ORGANISM Saccharum officinarum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Saccharum.
REFERENCE 1 (bases 1 to 683)
AUTHORS Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.
TITLE The libraries that made SUCEST
JOURNAL Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
COMMENT Contact: Arruda P
Centro de Biologia Molecular e Engenharia Genetica
Universidade Estadual de Campinas
Caixa Postal 6010, 13083-970, Campinas SP, Brazil
Tel: 55 19 3788 1137
Fax: 55 19 3788 1089
Email: parruda@unicamp.br
Clone distribution: clone distribution information can be found through the Brazilian Clone Collection Center (BCCC) at http://www.bcccenter.fcav.unesp.br

Plate: 093 row: A column: 04
Seq primer: T7 Promoter Primer.

FEATURES Location/Qualifiers
source 1. .683
/organism="Saccharum officinarum"
/mol_type="mRNA"
/db_xref="taxon:4547"
/clone="SCCCFL4093A04"
/lab_host="DH10B"
/clone_lib="FL4"
/note="Organ: Developed inflorescence and rachis (20cm-long); Vector: pSport1; Site_1: SalI; Site_2: NotI; An unidirectional cDNA library generated from [Developed inflorescence and rachis (20cm-long)]. cDNA was prepared from polyA+ mRNA using SuperScript Plasmid System Kit (Invitrogen). The double-strand cDNAs were fractionated in a sepharose CL-2B 40cm-columns and fragments sizing between 0.8 and 1.5 Kb were directionally cloned into the vector. Details of each source of RNA and library construction can be obtained at http://sucest.lad.ic.unicamp.br/public"

ORIGIN

Query Match 5.8%; Score 60; DB 14; Length 683;
Best Local Similarity 46.7%; Pred. No. 0.088;
Matches 189; Conservative 0; Mismatches 216; Indels 0; Gaps 0;

```
Qy        357 CAACCGGGAGCTCTGGAGGAAGGTGGGGACGGTCGTCTTCATGGTCCCCCTGGAGATGTT 416
              || || |   | | |  ||  ||||  | | |||  |  |  | || |     |
Db        254 CATCCTGCTCCACGCGCTGACCGTGGCCGCCGCCGTGGCCGGGTGCGCCGTCATCGCGGC 313

Qy        417 GGCCCTCAACCTGGGGGTCACCCGGCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGAA 476
              | ||  |   | | ||| |  | |  | |||||||||   | |   |  |||  | |
Db        314 GCCCGACCCGCCGCGGGGCCGCTGCTACACCGTCCACATGTCGGCCACCGTCGTCGTCTC 373

Qy        477 AAAGGGCCTGGTGGCCACCGACGTCCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCAT 536
              |     | ||  ||| ||| ||| ||   |    ||  || || ||||  ||   ||
Db        374 CATCCTGCAGGGCGCCGCCGCCGTGCTCGCCTTCTCCCGCACCGCGGAGTTCCTCTCCGA 433

Qy        537 CGGCACCCTTTGGGCCGTCCGGCTGAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACTA 596
              |||   |  | |  ||||||   |  |   ||   | ||  |||  | ||| || |
Db        434 CGGGCTCAGGTCGTACGTCCGCGAGGAGGACGGCGCCGTCATCCTCCGCATGGTCGGCGG 493

Qy        597 CATCTACCCCTGGAGGAACCTCGCCCTAGACATGGCCAACGGCGTGCTCTCCTTCAACTG 656
              ||  |  |   |   |  |  |      | ||   ||  |   ||  | |  |  | |
Db        494 GCTCGGCGTCGCCATCTTCTGCCTCGAGTGGGTCGCGCTCGCGCTCGCCTTCGTGCTCAG 553

Qy        657 GGTCAAGGCCTACCAGGACCACGGAATCCGCCCCACCCTGGACGTGCTGGTCCTCTGGGC 716
              |  | | | ||||   ||| | ||| | || |     |  || | ||    || |
Db        554 GTACTACGTCTACGTTGACAAGGGAGTGCGGCGGCAACCCGATGCGCCCCACCGCCCAAG 613

Qy        717 TCAGGGGAAAAGGGTGATGCCCAACACCAAGACCGTGGCCGTTGA 761
              || |||| || |   |  | ||  ||||  | ||  | || || |
Db        614 TCGGGGGGAAGGAACGGGGNCCCGCACCTGGCCCTGGCCCTTTCA 658
```

RESULT 2
CNS0091P/c
LOCUS CNS0091P 925 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL053013
VERSION AL053013.1 GI:4934461
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 925)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES Location/Qualifiers
source 1. .925
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR19D16"
/clone_lib="RPCI-98"
/note="end : TET3"

ORIGIN

Query Match 5.3%; Score 54.6; DB 29; Length 925;
Best Local Similarity 15.8%; Pred. No. 1.2;
Matches 63; Conservative 163; Mismatches 172; Indels 0; Gaps 0;

```
Qy        262 CCCCCCGTTGGGAATGTATGTCAAGCCGTCCAGGACGGGCCTCAGAAGCTTCTGGAACTC 321
              |:|::::  :::: |:  : :  :|:: :: :: :: ::: : :  |:  : :|  : |
Db        917 CSCSSBSCSSSSMSTSSSNSSBCSSCSSSBSSSSTSSMSSSSBSSSSGSSSSSSGTSSAC 858

Qy        322 CTCCAGGAGATTGCCCGCTCCACCATCCCCTACGGCAACCGGGAGCTCTGGAGGAAGGTG 381
              ::| |::    || : : |: ::  :||  |: : : :  :|    :  || |||  :::
Db        857 VKCNASSSCGCCGCGMABCCMCSSSSSCCGSASARGVKVRASGGAGKRGGGSGGASASHS 798

Qy        382 GGGACGGTCGTCTTCATGGTCCCCCTGGAGATGTTGGCCCTCAACCTGGGGGTCACCCGG 441
              :::||:: :: |  : ::  ::::  ::::: |   |  :  :   : ::::: :| ::::
Db        797 SSSACBSSSSSCSASCWSASSSSSASSSRSRSGGGAGGSGASSSRSSSSSSSASAGSVVS 738

Qy        442 CAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGAAAAAGGGCCTGGTGGCCACCGACGTC 501
              :|: ::: |: :::| ::|  ::  : ::  :|:|    ::::       :||  :  |: :
Db        737 SASSSSSSCSSSVSCSSVASSMSCSSBSSSSASASSSSSSSSASCASCSCCCTSWSCSCS 678

Qy        502 CTTCACCAAACCGTCAACGGGGAGCGCCGGGCCATCGGCACCCTTTGGGCCGTCCGGCTG 561
                 ::: |: ::: :  |:::::: ::: ::::| ::::  ::   |: | : :::|  :
Db        677 TSASMSAARSSSSSSSSCSSSSMSASSSASSSSASSSSSSGSSSSSGSACGBSMSSGGGS 618

Qy        562 AGGCCAGGGAAAGCCAGGCTCACCCTGGACGACTACATCTACCCCTGGAGGAACCTCGCC 621
              ::: :|   ::: ||  : :   ||: |   :  :  :  : :: :    | |:|:
Db        617 GSVSASSGMSSSVSSSGGRSSGSGGGGGVGGSGSSSSGSGSSGGSGSVCSCSSGCMCRCS 558

Qy        622 CTAGACATGGCCAACGGCGTGCTCTCCTTCAACTGGGT 659
              |  | |  :|:||: |:  |    ||    |  |
Db        557 CSSAAAAAASCVAASCGMMCGKSKSGCTNNNNNTNTNT 520
```

RESULT 3
CG352167
LOCUS CG352167 655 bp DNA linear GSS 26-AUG-2003
DEFINITION OG1BW12TV ZM_0.7_1.5_KB Zea mays genomic clone ZMMBMa0727A23, genomic survey sequence.
ACCESSION CG352167
VERSION CG352167.1 GI:34269433
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD

clade; Paniceae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 655)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T., Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
TITLE Consortium for Maize Genomics
JOURNAL Unpublished (2002)
COMMENT Other_GSSs: OG1BW12TH
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: sheared ends.
FEATURES Location/Qualifiers
source 1. .655
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBMa0727A23"
/clone_lib="ZM_0.7_1.5_KB"
/note="Vector: pBCSK-; Site_1: HincII; 0.7-1.5 kb methylation filtered genomic DNA library"
ORIGIN

Query Match 5.2%; Score 53.2; DB 29; Length 655;
Best Local Similarity 46.1%; Pred. No. 2.1;
Matches 178; Conservative 0; Mismatches 208; Indels 0; Gaps 0;

```
Qy        321 CCTCCAGGAGATTGCCCGCTCCACCATCCCCTACGGCAACCGGGAGCTCTGGAGGAAGGT 380
Db        145 CTTCGAAATGCTGTCGCCGTCCTTCACCGCGTACGGTACCTGCAACGCCGCGAACCGCGC 204

Qy        381 GGGGACGGTCGTCTTCATGGTCCCCCTGGAGATGTTGGCCCTCAACCTGGGGGTCACCCG 440
Db        205 GCTCACGGGCTGCCTCATCGGCGCCTGCGCGCTCTGCTGCTTCGTCCTCTGCTTCACCGA 264

Qy        441 GCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGAAAAAGGGCCTGGTGGCCACCGACGT 500
Db        265 CAGCTACCGCGACACCGCCTCGGGCGCGCTGCGGTACGGCTTCATCACACCCATCGGCCG 324

Qy        501 CCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCATCGGCACCCTTTGGGCCGTCCGGCT 560
Db        325 CCTGCTCCCCATCGAGGGTGGCGGCTCCAGCTCCGGCCCCGGCCCCGGCGCGCCGCCGCC 384

Qy        561 GAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACTACATCTACCCCTGGAGGAACCTCGC 620
Db        385 GCGGGACGACAGGTACAGGCTGACCGTGCGCGACGTCATGCACGGGCTGCTGTCTTTCGC 444

Qy        621 CCTAGACATGGCCAACGGCGTGCTCTCCTTCAACTGGGTCAAGGCCTACCAGGACCACGG 680
Db        445 GGTGTTCCTGGCCGTGGCCATGGTGGACCGCAACGTGGTGGCGTGCTTCTACCCCGTGGA 504

Qy        681 AATCCGCCCCACCCTGGACGTGCTGG 706
Db        505 GTCCGCGTCCACCATGCAGCTGCTGG 530
```

RESULT 4
CC609109/c
LOCUS CC609109 679 bp DNA linear GSS 18-JUN-2003
DEFINITION OGWDJ44TH ZM_0.7_1.5_KB Zea mays genomic clone ZMMBMa0544H16, genomic survey sequence.
ACCESSION CC609109
VERSION CC609109.1 GI:31970530
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 679)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T., Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
TITLE Consortium for Maize Genomics
JOURNAL Unpublished (2002)
COMMENT Other_GSSs: OGWDJ44TV
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TR
Class: sheared ends.
FEATURES Location/Qualifiers
source 1. .679
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBMa0544H16"
/clone_lib="ZM_0.7_1.5_KB"
/note="Vector: pBCSK-; Site_1: HincII; 0.7-1.5 kb methylation filtered genomic DNA library"
ORIGIN

Query Match 5.2%; Score 53.2; DB 29; Length 679;
Best Local Similarity 46.1%; Pred. No. 2.1;
Matches 178; Conservative 0; Mismatches 208; Indels 0; Gaps 0;

```
Qy        321 CCTCCAGGAGATTGCCCGCTCCACCATCCCCTACGGCAACCGGGAGCTCTGGAGGAAGGT 380
Db        622 CTTCGAAATGCTGTCGCCGTCCTTCACCGCGTACGGTACCTGCAACGCCGCGAACCGCGC 563

Qy        381 GGGGACGGTCGTCTTCATGGTCCCCCTGGAGATGTTGGCCCTCAACCTGGGGGTCACCCG 440
Db        562 GCTCACGGGCTGCCTCATCGGCGCCTGCGCGCTCTGCTGCTTCGTCCTCTGCTTCACCGA 503

Qy        441 GCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGAAAAAGGGCCTGGTGGCCACCGACGT 500
Db        502 CAGCTACCGCGACACCGCCACGGGCGCGCTGCGGTACGGCTTCATCACACCCAGCGGCCG 443

Qy        501 CCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCATCGGCACCCTTTGGGCCGTCCGGCT 560
Db        442 CCTGCTCCCCATCGAGGGTGGCGGCTCCAGCTCCGGCCCCGGCCCCGGCGCGCCGCCGCC 383

Qy        561 GAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACTACATCTACCCCTGGAGGAACCTCGC 620
Db        382 GCGGGACGACAGGTACAGGCTGACCGTGCGCGACGTCATGCACGGGCTGCTGTCGTTCGC 323

Qy        621 CCTAGACATGGCCAACGGCGTGCTCTCCTTCAACTGGGTCAAGGCCTACCAGGACCACGG 680
Db        322 GGTGTTCCTGGCCGTGGCCATGGTGGACCGCAACGTGGTGGCGTGCTTCTACCCCGTGGA 263

Qy        681 AATCCGCCCCACCCTGGACGTGCTGG 706
Db        262 GTCCGCGTCCACCAGGCAGCTGCTGG 237
```

RESULT 5
CC661222/c
LOCUS CC661222 773 bp DNA linear GSS 19-JUN-2003
DEFINITION OGLAU18TV ZM_0.7_1.5_KB Zea mays genomic clone ZMMBMa0312C11, genomic survey sequence.
ACCESSION CC661222
VERSION CC661222.1 GI:32064977
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD

clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 773)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T., Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
TITLE Consortium for Maize Genomics
JOURNAL Unpublished (2002)
COMMENT Other_GSSs: OGLAU18TH
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: sheared ends.
FEATURES Location/Qualifiers
source 1. .773
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBMa0312C11"
/clone_lib="ZM_0.7_1.5_KB"
/note="Vector: pBCSK-; Site_1: HincII; 0.7-1.5 kb methylation filtered genomic DNA library"
ORIGIN

Query Match 5.2%; Score 53.2; DB 29; Length 773;
Best Local Similarity 46.1%; Pred. No. 2.2;
Matches 178; Conservative 0; Mismatches 208; Indels 0; Gaps 0;

```
Qy        321 CCTCCAGGAGATTGCCCGCTCCACCATCCCCTACGGCAACCGGGAGCTCTGGAGGAAGGT 380
              | ||| |   | |  | |  |||  || | | ||||| | | |  |   |  ||     |
Db        395 CTTCGAAATGCTGTCGCCGTCCTTCACCGCGTACGGTACCTGCAACGCCGCGAACCGCGC 336

Qy        381 GGGGACGGTCGTCTTCATGGTCCCCCTGGAGATGTTGGCCCTCAACCTGGGGGTCACCCG 440
              |   |||| |  | |||| | | ||   | | | |    | ||  |||  |  |||||
Db        335 GCTCACGGGCTGCCTCATCGGCGCCTGCGCGCTCTGCTGCTTCGTCCTCTGCTTCACCGA 276

Qy        441 GCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGAAAAAGGGCCTGGTGGCCACCGACGT 500
                | | ||  ||   | |   |  ||   | |       | |     |||  ||  |
Db        275 CAGCTACCGCGACACCGCCACGGGCGCGCTGCGGTACGGCTTCATCACACCCAGCGGCCG 216

Qy        501 CCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCATCGGCACCCTTTGGGCCGTCCGGCT 560
              ||| | ||  | ||      || |   | |  ||  | |  ||   | ||     | ||
Db        215 CCTGCTCCCCATCGAGGGTGGCGGCTCCAGCTCCGGCCCCGGCCCCGGCGCGCCGCCGCC 156

Qy        561 GAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACTACATCTACCCCTGGAGGAACCTCGC 620
              | ||   | |   ||||||  ||| ||  ||||  |||  ||     |  |    ||||
Db        155 GCGGGACGACAGGTACAGGCTGACCGTGCGCGACGTCATGCACGGGCTGCTGTCGTTCGC 96

Qy        621 CCTAGACATGGCCAACGGCGTGCTCTCCTTCAACTGGGTCAAGGCCTACCAGGACCACGG 680
                |   | |||||   | | ||  |   ||||  |||   |  || | |   |   |
Db         95 GGTGTTCCTGGCCGTGGCCATGGTGGACCGCAACGTGGTGGCGTGCTTCTACCCCGTGGA 36

Qy        681 AATCCGCCCCACCCTGGACGTGCTGG 706
                |    |||||  | |  ||||||
Db         35 GTCCGCGTCCACCAGGCAGCTGCTGG 10
```

RESULT 6
BZ529007
LOCUS BZ529007 807 bp DNA linear GSS 16-DEC-2002
DEFINITION OGAIR91TC ZM2_0.7_1.5_KB Zea mays genomic clone ZMMBMa0070O14, genomic survey sequence.
ACCESSION BZ529007
VERSION BZ529007.1 GI:27070951
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD

clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 807)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T., Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
TITLE Consortium for Maize Genomics
JOURNAL Unpublished (2002)
COMMENT Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: sheared ends.
FEATURES Location/Qualifiers
source 1. .807
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBMa0070O14"
/clone_lib="ZM2_0.7_1.5_KB"
/note="Vector: pBCSK-; Site_1: HincII; 0.7-1.5 kb methylation filtered genomic DNA library"
ORIGIN

Query Match 5.2%; Score 53.2; DB 28; Length 807;
Best Local Similarity 46.1%; Pred. No. 2.2;
Matches 178; Conservative 0; Mismatches 208; Indels 0; Gaps 0;

```
Qy        321 CCTCCAGGAGATTGCCCGCTCCACCATCCCCTACGGCAACCGGGAGCTCTGGAGGAAGGT 380
              | ||| |   | |  | |  |||  || | | ||||| | | |  |   |  ||     |
Db        142 CTTCGAAATGCTGTCGCCGTCCTTCACCGCGTACGGTACCTGCAACGCCGCGAACCGCGC 201

Qy        381 GGGGACGGTCGTCTTCATGGTCCCCCTGGAGATGTTGGCCCTCAACCTGGGGGTCACCCG 440
              |   |||| |  | |||| | | ||   | | | |    | ||  |||  |  |||||
Db        202 GCTCACGGGCTGCCTCATCGGCGCCTGCGCGCTCTGCTGCTTCGTCCTCTGCTTCACCGA 261

Qy        441 GCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGAAAAAGGGCCTGGTGGCCACCGACGT 500
                | | ||  ||   | |   |  ||   | |       | |     |||  ||  |
Db        262 CAGCTACCGCGACACCGCCACGGGCGCGCTGCGGTACGGCTTCATCACACCCAGCGGCCG 321

Qy        501 CCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCATCGGCACCCTTTGGGCCGTCCGGCT 560
              ||| | ||  | ||      || |   | |  ||  | |  ||   | ||     | ||
Db        322 CCTGCTCCCCATCGAGGGTGGCGGCTCCAGCTCCGGCCCCGGCCCCGGCGCGCCGCCGCC 381

Qy        561 GAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACTACATCTACCCCTGGAGGAACCTCGC 620
              | ||   | |   ||||||  ||| ||  ||||  |||  ||     |  |    ||||
Db        382 GCGGGACGACAGGTACAGGCTGACCGTGCGCGACGTCATGCACGGGCTGCTGTCGTTCGC 441

Qy        621 CCTAGACATGGCCAACGGCGTGCTCTCCTTCAACTGGGTCAAGGCCTACCAGGACCACGG 680
                |   | |||||   | | ||  |   ||||  |||   |  || | |   |   |
Db        442 GGTGTTCCTGGCCGTGGCCATGGTGGACCGCAACGTGGTGGCGTGCTTCTACCCCGTGGA 501

Qy        681 AATCCGCCCCACCCTGGACGTGCTGG 706
                |    |||||  | |  ||||||
Db        502 GTCCGCGTCCACCAGGCAGCTGCTGG 527
```

RESULT 7
CG034209
LOCUS CG034209 815 bp DNA linear GSS 19-AUG-2003
DEFINITION PUJGY50TD ZM_0.6_1.0_KB Zea mays genomic clone ZMMBTa0666I03, genomic survey sequence.
ACCESSION CG034209
VERSION CG034209.1 GI:33906365
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.

```
REFERENCE   1  (bases 1 to 815)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
            Bennetzen,J.
  TITLE     Maize Genomics Consortium
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1..815
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBTa0666I03"
                     /clone_lib="ZM_0.6_1.0_KB"
                     /note="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
                     CoT selected genomic DNA library"
ORIGIN

  Query Match             5.2%;  Score 53.2;  DB 29;  Length 815;
  Best Local Similarity   46.1%;  Pred. No. 2.2;
  Matches  178;  Conservative    0;  Mismatches  208;  Indels    0;  Gaps    0;

Qy        321 CCTCCAGGAGATTGCCCGCTCCACCATCCCCTACGGCAACCGGGAGCTCTGGAGGAAGGT 380
Db        138 CTTCGAAATGCTGTCGCCGTCCTTCACCGCGTACGGTACCTGCAACGCCGCGAACCGCGC 197

Qy        381 GGGGACGGTCGTCTTCATGGTCCCCCTGGAGATGTTGGCCCTCAACCTGGGGGTCACCCG 440
Db        198 GCTCACGGGCTGCCTCATCGGCGCCTGCGCGCTCTGCTGCTTCGTCCTCTGCTTCACCGA 257

Qy        441 GCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGAAAAAGGGCCTGGTGGCCACCGACGT 500
Db        258 CAGCTACCGCGACACCGCCACGGGCGCGCTGCGGTACGGCTTCATCACACCCAGCGGCCG 317

Qy        501 CCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCATCGGCACCCTTTGGGCCGTCCGGCT 560
Db        318 CCTGCTCCCCATCGAGGGTGGCGGCTCCAGCTCCGGCCCCGGCCCCGGCGCGCCGCCGCC 377

Qy        561 GAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACTACATCTACCCCTGGAGGAACCTCGC 620
Db        378 GCGGGACGACAGGTACAGGCTGACCGTGCGCGACGTCATGCACGGGCTGCTGTCGTTCGC 437

Qy        621 CCTAGACATGGCCAACGGCGTGCTCTCCTTCAACTGGGTCAAGGCCTACCAGGACCACGG 680
Db        438 GGTGTTCCTGGCCGTGGCCATGGTGGACCGCAACGTGGTGGCGTGCTTCTACCCCGTGGA 497

Qy        681 AATCCGCCCCACCCTGGACGTGCTGG 706
Db        498 GTCCGCGTCCACCAGGCAGCTGCTGG 523


RESULT 8
CC661216
LOCUS       CC661216                 897 bp    DNA     linear   GSS 19-JUN-2003
DEFINITION  OGLAU18TH ZM_0.7_1.5_KB Zea mays genomic clone ZMMBMa0312C11,
            genomic survey sequence.
ACCESSION   CC661216
VERSION     CC661216.1  GI:32064967
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 897)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
            Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
  TITLE     Consortium for Maize Genomics
  JOURNAL   Unpublished (2002)
COMMENT     Other_GSSs: OGLAU18TV
            Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1..897
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBMa0312C11"
                     /clone_lib="ZM_0.7_1.5_KB"
                     /note="Vector: pBCSK-; Site_1: HincII; 0.7-1.5 kb
                     methylation filtered genomic DNA library"
ORIGIN

  Query Match             5.2%;  Score 53.2;  DB 29;  Length 897;
  Best Local Similarity   46.1%;  Pred. No. 2.3;
  Matches  178;  Conservative    0;  Mismatches  208;  Indels    0;  Gaps    0;

Qy        321 CCTCCAGGAGATTGCCCGCTCCACCATCCCCTACGGCAACCGGGAGCTCTGGAGGAAGGT 380
Db        412 CTTCGAAATGCTGTCGCCGTCCTTCACCGCGTACGGTACCTGCAACGCCGCGAACCGCGC 471

Qy        381 GGGGACGGTCGTCTTCATGGTCCCCCTGGAGATGTTGGCCCTCAACCTGGGGGTCACCCG 440
Db        472 GCTCACGGGCTGCCTCATCGGCGCCTGCGCGCTCTGCTGCTTCGTCCTCTGCTTCACCGA 531

Qy        441 GCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGAAAAAGGGCCTGGTGGCCACCGACGT 500
Db        532 CAGCTACCGCGACACCGCCACGGGCGCGCTGCGGTACGGCTTCATCACACCCAGCGGCCG 591

Qy        501 CCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCATCGGCACCCTTTGGGCCGTCCGGCT 560
Db        592 CCTGCTCCCCATCGAGGGTGGCGGCTCCAGCTCCGGCCCCGGCCCCGGCGCGCCGCCGCC 651

Qy        561 GAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACTACATCTACCCCTGGAGGAACCTCGC 620
Db        652 GCGGGACGACAGGTACAGGCTGACCGTGCGCGACGTCATGCACGGGCTGCTGTCGTTCGC 711

Qy        621 CCTAGACATGGCCAACGGCGTGCTCTCCTTCAACTGGGTCAAGGCCTACCAGGACCACGG 680
Db        712 GGTGTTCCTGGCCGTGGCCATGGTGGACCGCAACGTGGTGGCGTGCTTCTACCCCGTGGA 771

Qy        681 AATCCGCCCCACCCTGGACGTGCTGG 706
Db        772 GTCCGCGTCCACCAGGCAGCTGCTGG 797


RESULT 9
CNS0091P
LOCUS       CNS0091P                 925 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL053013
VERSION     AL053013.1  GI:4934461
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
```

```
REFERENCE   1  (bases 1 to 925)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .925
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR19D16"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN
```

```
  Query Match             5.0%;  Score 51.8;  DB 29;  Length 925;
  Best Local Similarity   13.7%;  Pred. No. 4.4;
  Matches   54;  Conservative  172;  Mismatches  167;  Indels    0;  Gaps    0;

Qy        403 CCCCTGGAGATGTTGGCCCTCAACCTGGGGGTCACCCGGCAGACCGTCCACGCCTGGAAG 462
              |:::  | :  :||:|:  |   :: : :| |: | : :|:: : |:: :| :::|
Db        532 CSMSMCGKKCGSTTBGSTTTTTTSSGSGYGKGCSSGSGBSCSCCSSCSCSSSSCSCCBCC 591

Qy        463 AAGGTCCTTGAGAAAAAGGGCCTGGTGGCCACCGACGTCCTTCACCAAACCGTCAACGGG 522
              : :::  :   : :: :: : || : ::|||:|  : :|   :::|:  : ::
Db        592 CCCSCSSYCCSSSBSSSKCSSTSBSCSCCCSSKSVCGTSCSSSSSCSSSSSSTSSSSTSS 651

Qy        523 GAGCGCCGGGCCATCGGCACCCTTTGGGCCGTCCGGCTGAGGCCAGGGAAAGCCAGGCTC 582
              : : ::::|::: :::: :  |  :|: ::: ||  |  |:  :::   : : ::: :
Db        652 STSKSSSSGSSSSSSSSYTTSKSTSASGSGSWSAGGGSGSTGSTSSSSSSSSTSTSSSSV 711

Qy        583 ACCCTGGACGACTACATCTACCCCTGGAGGAACCTCGCCCTAGACATGGCCAACGGCGTG 642
              : :::: ::    :     ::::|:: ::  :  ::::  : :  : |   |: :  :
Db        712 SSGSKSSTBSSGSBSSSGSSSSSSTSSBBSCTSTSSSSSSSYSSSTCSCCTCCCSYSYSS 771

Qy        643 CTCTCCTTCAACTGGGTCAAGGCCTACCAGGACCACGGAATCCGCCCCACCCTGGACGTG 702
              :|: :: |:: :|:|: :  :| :  :   : || |   :  | : ::::|:|:  || :
Db        772 STSSSSSTSWGSTSGSSSSSVGTSSSSDSTSTCCSCCCYMCTCCSTYBMBCYTSTSCGGS 831

Qy        703 CTGGTCCTCTGGGCTCAGGGGAAAAGGGTGATGCCCAACACCAAGACCGTGGCCGTTGAC 762
              : ::    :|: ||   |:::   ::|  :| :::: | :: :: ::: : ::: : :
Db        832 SSSSGKGGVTKCGCGGCGSSSTNGMBGTSSACSSSSSSCSSSSVSSSSKSSASSSSVSSS 891

Qy        763 CTGGGCCTCATCCTGGTCCTCCCCGAGGTGGAG 795
              :|::: :   :::: :  ::: : :| |:::
Db        892 GSSGVSSNSSSASKSSSSGSVSSGSGSGSGSVS 924
```

```
RESULT 10
CA230834
LOCUS       CA230834                649 bp    mRNA    linear   EST 25-SEP-2003
DEFINITION  SCJFFL3C07G11.g Saccharum officinarum FL3 Saccharum officinarum
            cDNA clone SCJFFL3C07G11 5', mRNA sequence.
ACCESSION   CA230834
VERSION     CA230834.1  GI:35294765
KEYWORDS    EST.
SOURCE      Saccharum officinarum
  ORGANISM  Saccharum officinarum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Saccharum.
REFERENCE   1  (bases 1 to 649)
  AUTHORS   Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.
  TITLE     The libraries that made SUCEST
  JOURNAL   Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
COMMENT     Contact: Arruda P
            Centro de Biologia Molecular e Engenharia Genetica
            Universidade Estadual de Campinas
            Caixa Postal 6010, 13083-970, Campinas SP, Brazil
            Tel: 55 19 3788 1137
            Fax: 55 19 3788 1089
            Email: parruda@unicamp.br
            Clone distribution: clone distribution information can be found
            through the Brazilian Clone Collection Center (BCCC) at
            http://www.bcccenter.fcav.unesp.br
            Plate: C07  row: G  column: 11
            Seq primer: T7 Promoter Primer.
FEATURES             Location/Qualifiers
     source          1. .649
                     /organism="Saccharum officinarum"
                     /mol_type="mRNA"
                     /db_xref="taxon:4547"
                     /clone="SCJFFL3C07G11"
                     /lab_host="DH10B"
                     /clone_lib="Saccharum officinarum FL3"
                     /note="Organ: Base of developing inflorescence (5cm-long);
                     Vector: pSport1; Site_1: SalI; Site_2: NotI; An
                     unidirectional cDNA library generated from [Base of
                     developing  inflorescence (5cm-long)]. cDNA was prepared
                     from polyA+ mRNA using  SuperScript Plasmid System Kit
                     (Invitrogen). The double-strand cDNAs were  fractionated
                     in a sepharose CL-2B 40cm-columns and fragments sizing
                     between 0.8 and 1.5 Kb were directionally cloned into the
                     vector. Details  of each source of RNA and library
                     construction can be obtained at
                     http://sucest.lad.ic.unicamp.br/public"
ORIGIN
```

```
  Query Match             5.0%;  Score 51.2;  DB 14;  Length 649;
  Best Local Similarity   43.9%;  Pred. No. 5.2;
  Matches  218;  Conservative    0;  Mismatches  279;  Indels    0;  Gaps    0;

Qy        329 AGATTGCCCGCTCCACCATCCCCTACGGCAACCGGGAGCTCTGGAGGAAGGTGGGGACGG 388
              |||| |    || |  |||| || |||| ||        ||     ||||  |||  ||
Db        114 AGATCGAGGACTGCTACATCGTCTCCGGCGACGACTGCGTCGCCGTGAAGAGCGGGTGGG 173

Qy        389 TCGTCTTCATGGTCCCCCTGGAGATGTTGGCCCTCAACCTGGGGGTCACCCGGCAGACCG 448
              ||  | |  ||   |  | ||| |  ||  || | |   |||  ||| ||||
Db        174 ACGAGTACGGCATCAGGTTCAACATGCCGAGCCAGCACATCGTCATCAGGAGGCTGACCT 233

Qy        449 TCCACGCCTGGAAGAAGGTCCTTGAGAAAAAGGGCCTGGTGGCCACCGACGTCCTTCACC 508
              |  | ||   | ||  | | |          ||||   | |   |||  || | | | |
Db        234 GCATCTCCCCCACGAGCGCCATGATCGCGCTGGGCAGCGAGATGTCCGGCGGCATCCGCG 293

Qy        509 AAACCGTCAACGGGGAGCGCCGGGCCATCGGCACCCTTTGGGCCGTCCGGCTGAGGCCAG 568
              |       | || |||   |  ||||||  |||    | |||||||  || | | | | |
Db        294 ACGTGCGCGCCGAGGACAACGTCGCCATCAACACGGAGTCGGCCGTCAGGGTCAAGTCCG 353

Qy        569 GGAAAGCCAGGCTCACCCTGGACGACTACATCTACCCCTGGAGGAACCTCGCCCTAGACA 628
              |    | |||  |  | | | |    ||||  |  |   |  |   |  |  ||    ||
Db        354 GCGCGGGGAGGGGCGGCTTTGTCAGGGACATTTTCGTGCGCGGCCTCAGCCTCCACACCA 413

Qy        629 TGGCCAACGGCGTGCTCTCCTTCAACTGGGTCAAGGCCTACCAGGACCACGGAATCCGCC 688
              ||      |  |        |  |     |  |    |||  ||| ||      |  ||
Db        414 TGAAGTGGGTGTTCTGGATGACCGGCAACTACGGGCAGCACCCCGACAACACGTCCAACC 473

Qy        689 CCACCCTGGACGTGCTGGTCCTCTGGGCTCAGGGGAAAAGGGTGATGCCCAACACCAAGA 748
              ||| |          ||||  | |   |   |   |||  |  |  | | | ||
Db        474 CCAACGCCTTGCCCGAGGTCACCGGCATCAACTACAGCGACGTGTTCGCTGAGAACGTGA 533
```

```
Qy          749 CCGTGGCCGTTGACCTGGGCCTCATCCTGGTCCTCCCCGAGGTGGAGCGTTCCAAACTCC 808
Db          534 CCATGGCCGGCAGGATGGAGGGCATCCCCAACGACCCCTACACCGGGATCTGCATATNCA 593

Qy          809 CGGCCCTTATCACCCTC 825
Db          594 ACGTCACTGCCCGCCTC 610
```

```
RESULT 11
CB870888
LOCUS       CB870888                 649 bp    mRNA    linear   EST 03-JUL-2003
DEFINITION  HC15J04w CH Hordeum vulgare cDNA clone HC15J04 5-PRIME, mRNA
            sequence.
ACCESSION   CB870888
VERSION     CB870888.1  GI:30072868
KEYWORDS    EST.
SOURCE      Hordeum vulgare
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 649)
  AUTHORS   Varshney,R.K., Zhang,H., Burton,R., Stein,N., Langridge,P. and
            Graner,A.
  TITLE     Barley ESTs from coleoptile tissue
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Stein Nils
            Molecular Markers Group, Department Genbank
            Institute of Plant Genetics and Crop Plant Research (IPK)
            Corrensstr. 3, 06466, Gatersleben, Germany
            Tel: 039482-5522
            Fax: 039482-5595
            Email: stein@ipk-gatersleben.de
            Insert Length: 649   Std Error: 0.00
            Plate: 15  row: J  column: 4
            Seq primer: T7.
FEATURES             Location/Qualifiers
     source          1..649
                     /organism="Hordeum vulgare"
                     /mol_type="mRNA"
                     /cultivar="Sloop"
                     /db_xref="GABI:549289"
                     /db_xref="taxon:4513"
                     /clone="HC15J04"
                     /tissue_type="coleoptile"
                     /dev_stage="coleoptile, 3 days old"
                     /lab_host="DH10B"
                     /clone_lib="CH"
                     /note="Vector: pSPORT; Site_1: SalI (5-end of cDNA);
                     Site_2: NotI (3-end of cDNA); Due to the cloning system
                     used blue/white selection for recombinants is not 100%
                     reliable. Average insert size is 1.3 kb."
ORIGIN

  Query Match             5.0%;  Score 51.2;  DB 14;  Length 649;
  Best Local Similarity   46.3%;  Pred. No. 5.2;
  Matches  167;  Conservative    0;  Mismatches  194;  Indels    0;  Gaps    0;

Qy          447 CGTCCACGCCTGGAAGAAGGTCCTTGAGAAAAAGGGCCTGGTGGCCACCGACGTCCTTCA 506
Db          227 CATCAACCGCGAGAACTACGAGCTGGGGCTGCCGTGCATCGAGAAGGCCGGCGTGGCGCA 286

Qy          507 CCAAACCGTCAACGGGGAGCGCCGGGCCATCGGCACCCTTTGGGCCGTCCGGCTGAGGCC 566
Db          287 CAAGATCGACTTCCGCGAGGGCCCGGCGCTCCCCGTCCTGGACGCCCTCCTCGAGGACGA 346

Qy          567 AGGGAAAGCCAGGCTCACCCTGGACGACTACATCTACCCCTGGAGGAACCTCGCCCTAGA 626
Db          347 GGCCAACCACGGCACCTTCGACTTCGTCTTCGTGGACGCCGACAAGGACAACTACCTCAA 406

Qy          627 CATGGCCAACGGCGTGCTCTCCTTCAACTGGGTCAAGGCCTACCAGGACCACGGAATCCG 686
Db          407 CTACCACGAGCGCCTCATGAAGCTCGTCAAGGTCGGCGGCCTCCTCGGCTACGACAACAC 466

Qy          687 CCCCACCCTGGACGTGCTGGTCCTCTGGGCTCAGGGGAAAAGGGTGATGCCCAACACCAA 746
Db          467 CCTCTGGAACGGCTCCGTCGTGCTCCCCGCCGACGCCCCCATGCGCAAGTACATCCGCTA 526

Qy          747 GACCGTGGCCGTTGACCTGGGCCTCATCCTGGTCCTCCCCGAGGTGGAGCGTTCCAAACT 806
Db          527 CTACCGCGACTTCGTCCTCGACCTCAACAAGGCCCTCGCCGNCGACCAGCGCGTCGAGAT 586

Qy          807 C 807
Db          587 C 587
```

```
RESULT 12
CB874255/c
LOCUS       CB874255                 594 bp    mRNA    linear   EST 03-JUL-2003
DEFINITION  HC15J04y CH Hordeum vulgare cDNA clone HC15J04 3-PRIME, mRNA
            sequence.
ACCESSION   CB874255
VERSION     CB874255.1  GI:30076241
KEYWORDS    EST.
SOURCE      Hordeum vulgare
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 594)
  AUTHORS   Varshney,R.K., Zhang,H., Burton,R., Stein,N., Langridge,P. and
            Graner,A.
  TITLE     Barley ESTs from coleoptile tissue
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Stein Nils
            Molecular Markers Group, Department Genbank
            Institute of Plant Genetics and Crop Plant Research (IPK)
            Corrensstr. 3, 06466, Gatersleben, Germany
            Tel: 039482-5522
            Fax: 039482-5595
            Email: stein@ipk-gatersleben.de
            Insert Length: 594   Std Error: 0.00
            Plate: 15  row: J  column: 4
            Seq primer: SP6.
FEATURES             Location/Qualifiers
     source          1..594
                     /organism="Hordeum vulgare"
                     /mol_type="mRNA"
                     /cultivar="Sloop"
                     /db_xref="GABI:552657"
                     /db_xref="taxon:4513"
                     /clone="HC15J04"
                     /tissue_type="coleoptile"
                     /dev_stage="coleoptile, 3 days old"
                     /lab_host="DH10B"
                     /clone_lib="CH"
                     /note="Vector: pSPORT; Site_1: SalI (5-end of cDNA);
                     Site_2: NotI (3-end of cDNA); Due to the cloning system
                     used blue/white selection for recombinants is not 100%
                     reliable. Average insert size is 1.3 kb."
ORIGIN

  Query Match             4.9%;  Score 50.6;  DB 14;  Length 594;
  Best Local Similarity   46.3%;  Pred. No. 6.7;
  Matches  167;  Conservative    0;  Mismatches  194;  Indels    0;  Gaps    0;

Qy          447 CGTCCACGCCTGGAAGAAGGTCCTTGAGAAAAAGGGCCTGGTGGCCACCGACGTCCTTCA 506
Db          588 CATCAACCGCGAGAACTACGAGCTGGGGCTGCCGTGCATCGAGAAGGCCGGCGTGGCGCA 529

Qy          507 CCAAACCGTCAACGGGGAGCGCCGGGCCATCGGCACCCTTTGGGCCGTCCGGCTGAGGCC 566
Db          528 CAAGATCGACTTCCGCGAGGGCCCGGCGCTCCCCGTCCTGGACGCCCTCCTCGAGGACGA 469
```

```
Qy        567 AGGGAAAGCCAGGCTCACCCTGGACGACTACATCTACCCCTGGAGGAACCTCGCCCTAGA 626
Db        468 GGCCAACCACGGCACCTTCGACTTCGTCTTCGTGGACGCCGACAAGGACAACTACCTCAA 409

Qy        627 CATGGCCAACGGCGTGCTCTCCTTCAACTGGGTCAAGGCCTACCAGGACCACGGAATCCG 686
Db        408 CTACCACGAGCGCCTCATGAAGCTCGTCAAGGTCGGCGGCCTCCTCGGCTACGACAACAC 349

Qy        687 CCCCACCCTGGACGTGCTGGTCCTCTGGGCTCAGGGGAAAAGGGTGATGCCCAACACCAA 746
Db        348 CCTCTGGAACGGCTCCGTCGTGCTCCCCGCCGACGCCCCCATGCGCAAGTACATCCGCTA 289

Qy        747 GACCGTGGCCGTTGACCTGGGCCTCATCCTGGTCCTCCCCGAGGTGGAGCGTTCCAAACT 806
Db        288 CTACCGCGACTTCGTCCTCGACCTCAACAAGGCCCTCGCCGCCGACCAGCGCGTCGAGAT 229

Qy        807 C 807
Db        228 C 228
```

RESULT 13
BI960118
LOCUS BI960118 637 bp mRNA linear EST 22-OCT-2001
DEFINITION HVSMEn0023E18f Hordeum vulgare rachis EST library HVcDNA0015 (normal) Hordeum vulgare subsp. vulgare cDNA clone HVSMEn0023E18f, mRNA sequence.
ACCESSION BI960118
VERSION BI960118.1 GI:16311373
KEYWORDS EST.
SOURCE Hordeum vulgare subsp. vulgare
ORGANISM Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 637)
AUTHORS Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Chin,A., Begum,D., Frisch,D., Atkins,M., Yu,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Oates,R. and Main,D.
TITLE Development of a genetically and physically anchored EST resource for barley genomics: Morex rachis cDNA library
JOURNAL Unpublished (2001)
COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hq bases = 579
Seq primer: AATTAACCCTCACTAAAGGG
High quality sequence start: 5
High quality sequence stop: 633.
FEATURES Location/Qualifiers
source 1. .637
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="Morex"
/sub_species="vulgare"
/db_xref="taxon:112509"
/clone="HVSMEn0023E18f"
/tissue_type="Rachis"
/lab_host="TJC121"
/clone_lib="Hordeum vulgare rachis EST library HVcDNA0015 (normal)"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; Plants were grown at Washington State University, Pullman, WA in a greenhouse, the rachises were excised and frozen in liquid nitrogen (Kleinhofs lab). In the TJ Close lab at the University of California, Riverside total RNA was prepared, poly(A) was purified, one primary unamplified cDNA library was made, and 1 million pfu were in vivo excised to give pBluescript SK(-) cDNA phagemids (Chin). Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see http://www.genome.clemson.edu/projects/barley. To order this clone see http://www.genome.clemson.edu/orders Also see Close TJ, Wing R, Kleinhofs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"
ORIGIN

Query Match 4.9%; Score 50.6; DB 12; Length 637;
Best Local Similarity 46.3%; Pred. No. 6.8;
Matches 167; Conservative 0; Mismatches 194; Indels 0; Gaps 0;

```
Qy        447 CGTCCACGCCTGGAAGAAGGTCCTTGAGAAAAAGGGCCTGGTGGCCACCGACGTCCTTCA 506
Db        133 CATCAACCGCGAGAACTACGAGCTGGGGCTGCCGTGCATCGAGAAGGCCGGCGTGGCGCA 192

Qy        507 CCAAACCGTCAACGGGGAGCGCCGGGCCATCGGCACCCTTTGGGCCGTCCGGCTGAGGCC 566
Db        193 CAAGATCGACTTCCGCGAGGGCCCGGCGCTCCCCGTCCTGGACGCCCTCCTCGAGGACGA 252

Qy        567 AGGGAAAGCCAGGCTCACCCTGGACGACTACATCTACCCCTGGAGGAACCTCGCCCTAGA 626
Db        253 GGCCAACCACGGCACCTTCGACTTCGTCTTCGTGGACGCCGACAAGGACAACTACCTCAA 312

Qy        627 CATGGCCAACGGCGTGCTCTCCTTCAACTGGGTCAAGGCCTACCAGGACCACGGAATCCG 686
Db        313 CTACCACGAGCGCCTCATGAAGCTCGTCAAGGTCGGCGGCCTCCTCGGCTACGACAACAC 372

Qy        687 CCCCACCCTGGACGTGCTGGTCCTCTGGGCTCAGGGGAAAAGGGTGATGCCCAACACCAA 746
Db        373 CCTCTGGAACGGCTCCGTCGTGCTCCCCGCCGACGCCCCCATGCGCAAGTACATCCGCTA 432

Qy        747 GACCGTGGCCGTTGACCTGGGCCTCATCCTGGTCCTCCCCGAGGTGGAGCGTTCCAAACT 806
Db        433 CTACCGCGACTTCGTCCTCGACCTCAACAAGGCCCTCGCCGCCGACCAGCGCGTCGAGAT 492

Qy        807 C 807
Db        493 C 493
```

RESULT 14
BQ762267
LOCUS BQ762267 664 bp mRNA linear EST 26-JUL-2002
DEFINITION EBro01_SQ005_F11_R root, 3 week, hydroponic grown, no treatment, cv Optic, EBro01 Hordeum vulgare subsp. vulgare cDNA clone EBro01_SQ005_F11 5', mRNA sequence.
ACCESSION BQ762267
VERSION BQ762267.1 GI:21970739
KEYWORDS EST.
SOURCE Hordeum vulgare subsp. vulgare
ORGANISM Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 664)
AUTHORS Hedley,P., Liu,H., Caldwell,D., McCallum,N., Mudie,S., Cardle,L., Ramsay,L., Machray,G., Marshall,D.F.M. and Waugh,R.
TITLE Development of Barley Transcriptome Resources
JOURNAL Unpublished (2001)
COMMENT Contact: Waugh R, Marshall DF
Genome Dynamics/Computational Biology
Scottish Crop Research Institute
Invergowrie, Dundee, DD2 5DA, Scotland, UK

```
                Tel: 00 44 1382 562731
                Fax: 00 44 1382 562426
                Email: est@scri.sari.ac.uk.
FEATURES             Location/Qualifiers
     source          1..664
                     /organism="Hordeum vulgare subsp. vulgare"
                     /mol_type="mRNA"
                     /cultivar="Optic"
                     /sub_species="vulgare"
                     /db_xref="taxon:112509"
                     /clone="EBro01_SQ005_F11"
                     /tissue_type="root"
                     /dev_stage="3 week"
                     /lab_host="DH10B"
                     /clone_lib="root, 3 week, hydroponic grown, no treatment,
                     cv Optic, EBro01"
                     /note="Vector: pSPORT1; Site_1: Sal I; Site_2: Not I;
                     Non-normalised library, directionally cloned into pSPORT1.
                     Derived from roots of 3 week old hydroponically grown
                     unstressed barley plants.  Developed as part of the barley
                     transcriptome resources of BBSRC/SEERAD funded cereal IGF
                     (Investigating Gene Function) project."
ORIGIN

  Query Match             4.9%;  Score 50.6;  DB 13;  Length 664;
  Best Local Similarity   46.3%;  Pred. No. 6.9;
  Matches  167;  Conservative    0;  Mismatches  194;  Indels    0;  Gaps    0;

Qy        447 CGTCCACGCCTGGAAGAAGGTCCTTGAGAAAAAGGGCCTGGTGGCCACCGACGTCCTTCA 506
Db        140 CATCAACCGCGAGAACTACGAGCTGGGGCTGCCGTGCATCGAGAAGGCCGGCGTGGCGCA 199

Qy        507 CCAAACCGTCAACGGGGAGCGCCGGGCCATCGGCACCCTTTGGGCCGTCCGGCTGAGGCC 566
Db        200 CAAGATCGACTTCCGCGAGGGCCCGGCGCTCCCCGTCCTGGACGCCCTCCTCGAGGACGA 259

Qy        567 AGGGAAAGCCAGGCTCACCCTGGACGACTACATCTACCCCTGGAGGAACCTCGCCCTAGA 626
Db        260 GGCCAACCACGGCACCTTCGACTTCGTCTTCGTGGACGCCGACAAGGACAACTACCTCAA 319

Qy        627 CATGGCCAACGGCGTGCTCTCCTTCAACTGGGTCAAGGCCTACCAGGACCACGGAATCCG 686
Db        320 CTACCACGAGCGCCTCATGAAGCTCGTCAAGGTCGGCGGCCTCCTCGGCTACGACAACAC 379

Qy        687 CCCCACCCTGGACGTGCTGGTCCTCTGGGCTCAGGGGAAAAGGGTGATGCCCAACACCAA 746
Db        380 CCTCTGGAACGGCTCCGTCGTGCTCCCCGCCGACGCCCCCATGCGCAAGTACATCCGCTA 439

Qy        747 GACCGTGGCCGTTGACCTGGGCCTCATCCTGGTCCTCCCCGAGGTGGAGCGTTCCAAACT 806
Db        440 CTACCGCGACTTCGTCCTCGACCTCAACAAGGCCCTCGCCGCCGACCAGCGCGTCGAGAT 499

Qy        807 C 807
Db        500 C 500


RESULT 15
CA275829
LOCUS       CA275829                 973 bp    mRNA    linear   EST 26-SEP-2003
DEFINITION  SCCCSD1091C07.g SD1 Saccharum officinarum cDNA clone SCCCSD1091C07
            5', mRNA sequence.
ACCESSION   CA275829
VERSION     CA275829.1  GI:35994669
KEYWORDS    EST.
SOURCE      Saccharum officinarum
  ORGANISM  Saccharum officinarum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Saccharum.
REFERENCE   1  (bases 1 to 973)
  AUTHORS   Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.
  TITLE     The libraries that made SUCEST
  JOURNAL   Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
COMMENT     Contact: Arruda P
            Centro de Biologia Molecular e Engenharia Genetica
            Universidade Estadual de Campinas
            Caixa Postal 6010, 13083-970, Campinas SP, Brazil
            Tel: 55 19 3788 1137
            Fax: 55 19 3788 1089
            Email: parruda@unicamp.br
            Clone distribution: clone distribution information can be found
            through the Brazilian Clone Collection Center (BCCC) at
            http://www.bcccenter.fcav.unesp.br
            Plate: 091  row: C  column: 07
            Seq primer: T7 Promoter Primer.
FEATURES             Location/Qualifiers
     source          1..973
                     /organism="Saccharum officinarum"
                     /mol_type="mRNA"
                     /db_xref="taxon:4547"
                     /clone="SCCCSD1091C07"
                     /lab_host="DH10B"
                     /clone_lib="SD1"
                     /note="Organ: Developing seeds (large insert library);
                     Vector: pSport1; Site_1: SalI; Site_2: NotI; An
                     unidirectional cDNA library generated from [Developing
                     seeds (large  insert library)]. cDNA was prepared from
                     polyA+ mRNA using SuperScript  Plasmid System Kit
                     (Invitrogen). The double-strand cDNAs were  fractionated
                     in a sepharose CL-2B 40cm-columns and fragments sizing
                     between 0.8 and 1.5 Kb were directionally cloned into the
                     vector. Details  of each source of RNA and library
                     construction can be obtained at
                     http://sucest.lad.ic.unicamp.br/public"
ORIGIN

  Query Match             4.9%;  Score 50.6;  DB 14;  Length 973;
  Best Local Similarity   46.9%;  Pred. No. 7.8;
  Matches  158;  Conservative    0;  Mismatches  179;  Indels    0;  Gaps    0;

Qy        418 GCCCTCAACCTGGGGGTCACCCGGCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGAAA 477
Db         26 GCGCACGCCGCCGCCGTGCACCGCGAGGACGCCGACGGCCGCCTCGGCGTCAGCCTCAGC 85

Qy        478 AAGGGCCTGGTGGCCACCGACGTCCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCATC 537
Db         86 AAGGCGCTGCCCAAGGCCGCCGTGCGCGCCTTCACCGAGAACCTGCAGCGCCTGGCCCCG 145

Qy        538 GGCACCCTTTGGGCCGTCCGGCTGAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACTAC 597
Db        146 CGCATCCTCCCCGCCCGCGAGCTCGCGCGCTTCGCGTCCCGCCTCCTCCTCCGCAAGCTC 205

Qy        598 ATCTACCCCTGGAGGAACCTCGCCCTAGACATGGCCAACGGCGTGCTCTCCTTCAACTGG 657
Db        206 CTGCTGGCACCCCGCAACAACAGCAAGGCGGAGGGCGGCGCCAGCAGCCCCAAGATCGAC 265

Qy        658 GTCAAGGCCTACCAGGACCACGGAATCCGCCCCACCCTGGACGTGCTGGTCCTCTGGGCT 717
Db        266 TTCAAGACCGGCGTGGACCACTTCTGCCTCCACCCCGGCGGGACGGCGGTGATCGAGGCG 325

Qy        718 CAGGGGAAAAGGGTGATGCCCAACACCAAGACCGTGG 754
Db        326 GTGAGGAAGAGCCTGGGCCTCACCGCCCACGACGTGG 362


Search completed: March 18, 2004, 05:45:28
Job time : 1667.36 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: March 17, 2004, 20:21:47 ; Search time 14869.6 Seconds
(without alignments)
17049.054 Million cell updates/sec

Title: US-09-664-186-6
Perfect score: 5849
Sequence: 1 tctagaaggtcagggtggac.....tgtagaggaccagaagaccc 5849

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

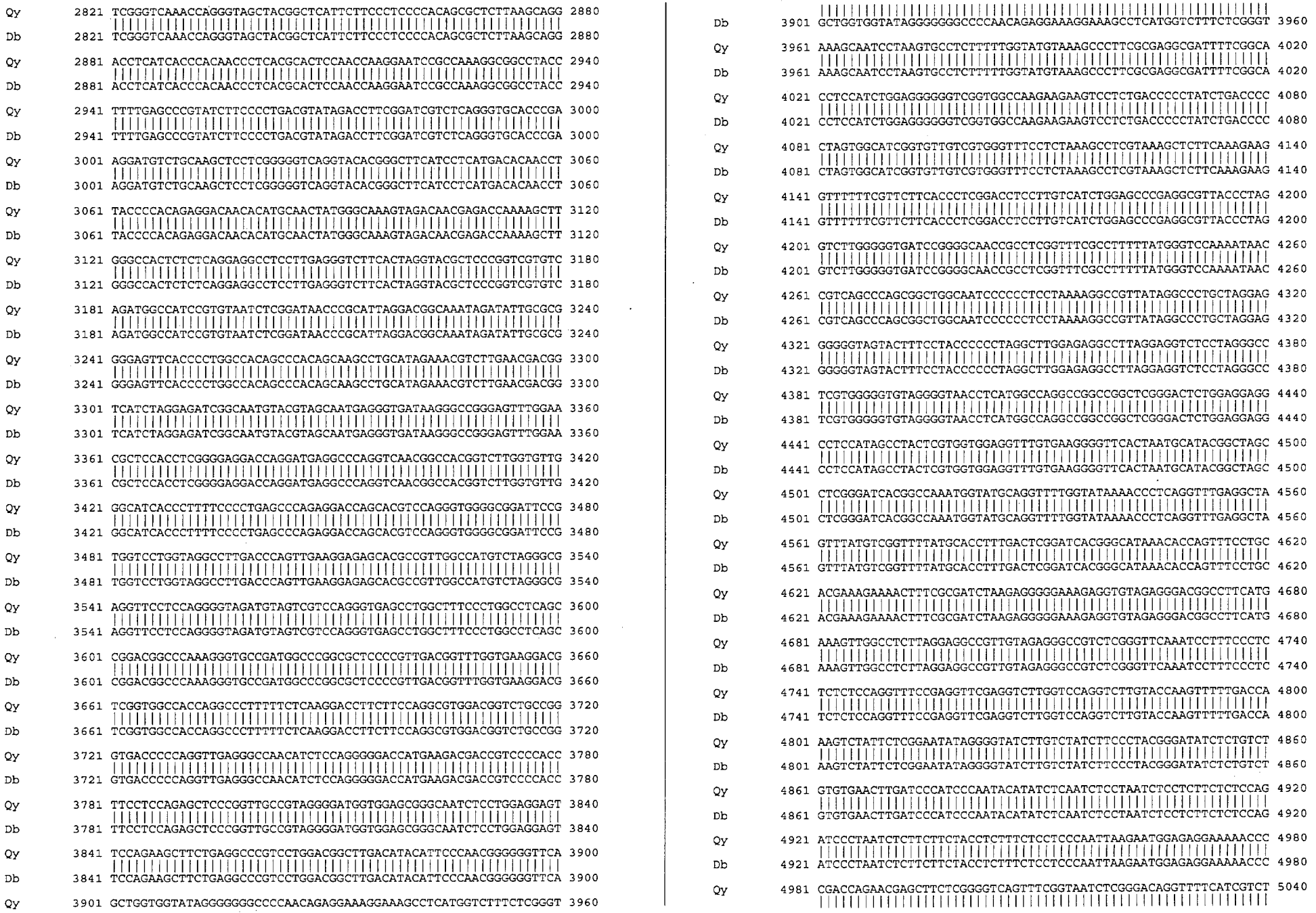
Result No.		Score	% Query Match	Length	DB	ID	Description
1		5849	100.0	5849	1	AF013571	AF013571 Thermus s
2		5849	100.0	5849	6	AR139426	AR139426 Sequence
3	c	1600	27.4	1600	1	U89376	U89376 Thermus sp.
4	c	1242	21.2	1242	6	AR031809	AR031809 Sequence
5	c	1026	17.5	1026	6	AR139424	AR139424 Sequence
6		999	17.1	999	6	AR031810	AR031810 Sequence
7		187.2	3.2	4938	1	AF328924	AF328924 Helicobac
8		185.6	3.2	11131	1	AE001444	AE001444 Helicobac
9		184	3.1	4931	1	AF328909	AF328909 Helicobac
10		182.4	3.1	3172	1	AF328912	AF328912 Helicobac
11		180.8	3.1	3261	1	AF328911	AF328911 Helicobac
12		180.8	3.1	5001	1	AF328915	AF328915 Helicobac
13	c	180	3.1	180	6	AR139425	AR139425 Sequence
14		179.2	3.1	3184	1	AF328910	AF328910 Helicobac
15		179.2	3.1	4931	1	AF328916	AF328916 Helicobac
16	c	108	1.8	1556	1	ECECAIMG	X17111 Enterobacte
17	c	77	1.3	7218	6	I66494	I66494 Sequence 14
18	c	61	1.0	3141	1	AF051092	AF051092 Xanthomon
19	c	59.8	1.0	303040	1	AP000991	AP000991 Thermopla
20	c	59.8	1.0	308050	1	SCO939124	AL939124 Streptomy
21	c	59	1.0	2000	6	AX655393	AX655393 Sequence
22		58.8	1.0	125020	9	AF429315	AF429315 Homo sapi
23		57.6	1.0	2000	6	AX655393	AX655393 Sequence
24	c	57.2	1.0	317022	1	AE017148	AE017148 Helicobac
25	c	56.8	1.0	125020	9	AF429315	AF429315 Homo sapi
26		56.8	1.0	300800	1	SCO939112	AL939112 Streptomy
27	c	56.4	1.0	2030	6	AX653795	AX653795 Sequence
28	c	56.4	1.0	2031	6	AX755741	AX755741 Sequence
29		56.4	1.0	7218	6	I66494	I66494 Sequence 14
30	c	55.4	0.9	2028	6	AX660328	AX660328 Sequence
31	c	55.4	0.9	2028	6	AX755777	AX755777 Sequence
32	c	54.8	0.9	2028	6	AX653794	AX653794 Sequence
33	c	54.8	0.9	2060	6	AX654685	AX654685 Sequence
34	c	54.8	0.9	2273	8	AK106696	AK106696 Oryza sat
35		54.8	0.9	137651	8	AC130600	AC130600 Oryza sat
36		54.8	0.9	177374	8	AC135429	AC135429 Oryza sat
37	c	54.6	0.9	202921	2	AC119314	AC119314 Rattus no
38	c	54.6	0.9	299050	1	SCO939119	AL939119 Streptomy
39		53.8	0.9	299425	1	AP005037	AP005037 Streptomy
40	c	53.6	0.9	303550	1	SCO939118	AL939118 Streptomy
41	c	53.2	0.9	1755	6	AX654077	AX654077 Sequence
42	c	53.2	0.9	1965	6	AX659936	AX659936 Sequence
43	c	53.2	0.9	1965	6	AX755739	AX755739 Sequence
44	c	53.2	0.9	2303	8	AK072823	AK072823 Oryza sat
45	c	53.2	0.9	100800	8	AC105260	AC105260 Oryza sat

ALIGNMENTS

RESULT 1
AF013571
LOCUS AF013571 5849 bp DNA linear BCT 03-JAN-1998
DEFINITION Thermus sp. YS45 plasmid pTsp45s: Tsp45I restriction endonuclease (tsp45IR), Tsp45I methyltransferase (tsp45IM), and replication protein (repT) genes, complete cds.
ACCESSION AF013571
VERSION AF013571.1 GI:2735918
KEYWORDS .
SOURCE Thermus sp. YS45
ORGANISM Thermus sp. YS45
Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae; Thermus.
REFERENCE 1 (bases 1 to 5849)
AUTHORS Wayne,J. and Xu,S.Y.

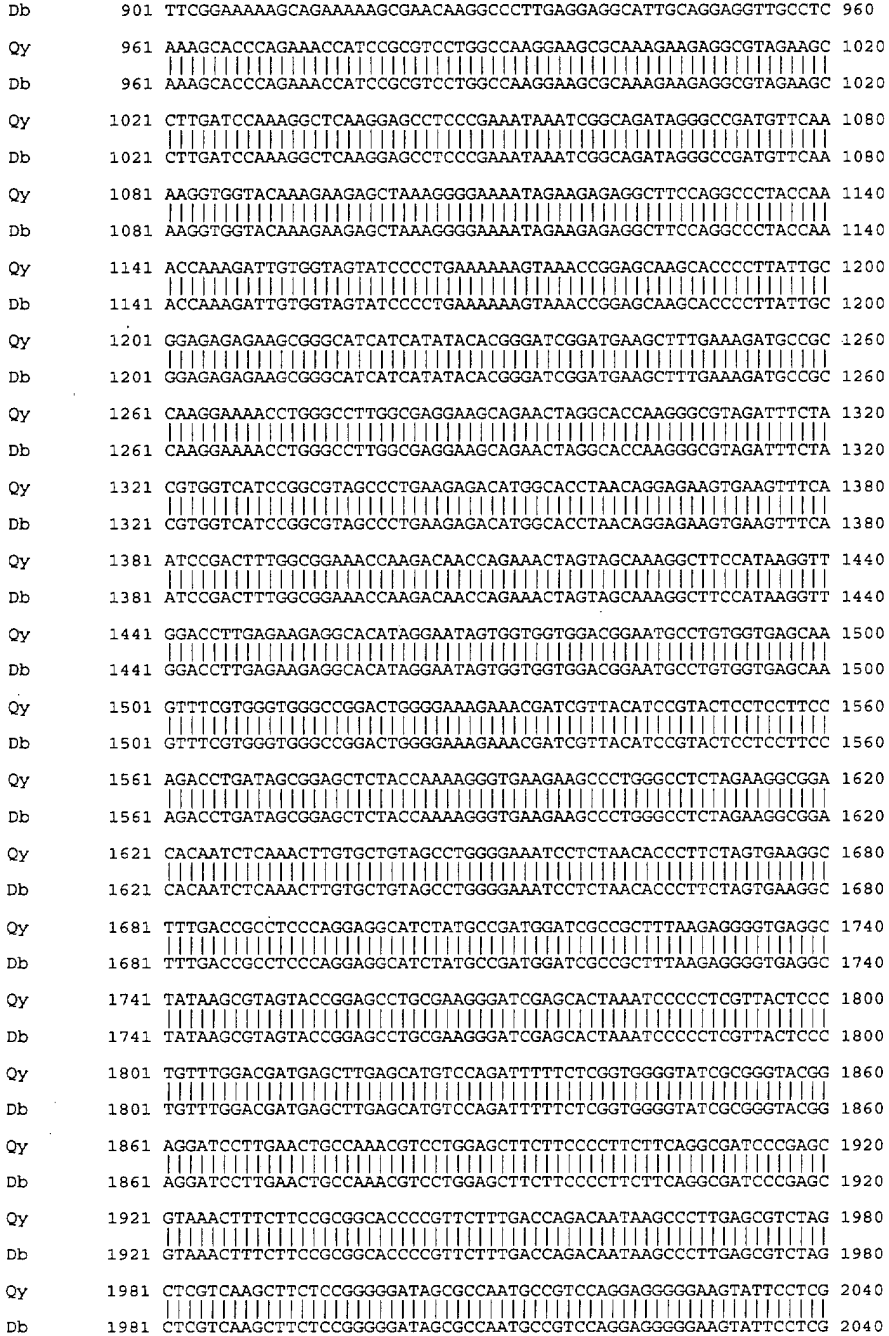
```
  TITLE     Identification of a thermophilic plasmid origin and its cloning
            within a new Thermus-E. coli shuttle vector
  JOURNAL   Gene 195 (2), 321-328 (1997)
  MEDLINE   97449309
   PUBMED   9305778
REFERENCE   2  (bases 1 to 5849)
  AUTHORS   Wayne,J., Holden,M. and Xu,S.Y.
  TITLE     The Tsp45I restriction-modification system is plasmid-borne within
            its thermophilic host
  JOURNAL   Gene 202 (1-2), 83-88 (1997)
  MEDLINE   98087421
   PUBMED   9427549
REFERENCE   3  (bases 1 to 5849)
  AUTHORS   Wayne,J., Holden,M. and Xu,S.-Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-1997) Research, New England Biolabs, 32 Tozer
            Road, Beverly, MA 01915, USA
FEATURES             Location/Qualifiers
     source          1. .5849
                     /organism="Thermus sp. YS45"
                     /mol_type="genomic DNA"
                     /strain="YS45"
                     /db_xref="taxon:64227"
                     /plasmid="pTsp45s"
     gene            <602. .>1612
                     /gene="tsp45IR"
     RBS             602. .606
                     /gene="tsp45IR"
     CDS             623. .1612
                     /gene="tsp45IR"
                     /note="R Tsp45I"
                     /codon_start=1
                     /transl_table=11
                     /product="Tsp45I restriction endonuclease"
                     /protein_id="AAC46043.1"
                     /db_xref="GI:2735919"
                     /translation="MAEWNVWTQRSVELLEKGYLDKLLQVYKGESGSSRSVPEEVEEK
                     LREAYKAYEGRQDSPEAETKLVEAVLNARKKVERSPFNHPYLPLVYYLVSEKAEKANK
                     ALEEALQEVASKHPETIRVLAKEAQRRGVEALIQRLKEPPEINRQIGPMFKRWYKEEL
                     KGKIEERLPGPTKPKIVVVSPEKSKPEQAPLIAEREAGIIIYTGSDEALKDAAKENLG
                     LGEEAELGTKGVDFYVVIRRSPEETWHLTGEVKFQSDFGGNQDNQKLVAKASIRLDLE
                     KRHIGIVVVDGMPVVSKFRGWAGLGKETIVTSVLLLPDLIAELYQKGEEALGL"
     gene            complement(<1609. .>2860)
                     /gene="tsp45IM"
     CDS             complement(1609. .2850)
                     /gene="tsp45IM"
                     /note="M Tsp45I"
                     /codon_start=1
                     /transl_table=11
                     /product="Tsp45I methyltransferase"
                     /protein_id="AAC46044.1"
                     /db_xref="GI:2735920"
                     /translation="MSRSYPGLTRKAPLKASKTSEPSPFRLVYPGKRDEKEILDQPTP
                     QLVLRKETLLFLGGNAPLFEIDPIGTYFLGENGQVLRWMLREPGGYAGKVQLVYIDPP
                     YGTGQQFLVGGDETDRVATVSQPKNGQLGYDDTLDGPQFVEFLRERLILLLRELMADSG
                     LIFVHIDEKYGFEVKLILDEVFGRRNFVNHIARIASNPKNFSRKAFGSQKDMILVYSK
                     TRDYVWNESASPYSEEEIARLFPFVDENGERYTTNPLHAPGETKDGPTGRPWRGILPP
                     PGRHWRYPPEKLDELDAQGLIVWSKNGVPRKKVYARDRLKKGKKLQDVWQFKDPPYPR
                     YPTEKNLDMLKLIVQTGSNEGDLVLDPFAGSGTTLIASPLLKRRSIGIDASWEAVKAF
                     TRRVLEDFPRLQHKFEIVSAF"
     rep_origin      2164. .4500
                     /note="required for plasmid replication (putative)"
     RBS             complement(2856. .2860)
                     /gene="tsp45IM"
     gene            complement(<3132. .>4169)
                     /gene="repT"
     CDS             complement(3132. .4157)
                     /gene="repT"
                     /function="required for plasmid replication"
                     /note="RepT"
                     /codon_start=1
                     /transl_table=11
                     /product="replication protein"
                     /protein_id="AAC46045.1"
                     /db_xref="GI:2735921"
                     /translation="MKNEKTFFEELYEALEETHDNTDATRGSDRGSEDFFLATDPPPD
                     GGAENRLAKGFTYQKEALRIALPEKDHEAFLSSVGAPPIPPAEPPVGNVCQAVQDGPQ
                     KLLELLQEIARSTIPYGNRELWRKVGTVVFMVPLEMLALNLGVTRQTVHAWKKVLEKK
                     GLVATDVLHQTVNGERRAIGTLWAVRLRPGKARLTLDDYIYPWRNLALDMANGVLSFN
                     WVKAYQDHGIRPTLDVLVLWAQGKRVMPNTKTVAVDLGLILVLPEVERSKLPALITLI
                     ATYIADLLDDRRSRRFYAGLLWAVARGELPAQYLFAVLMRVIRDYTDGHLTRPGAYLV
                     KTLKEAS"
     misc_feature    complement(3200. .3208)
                     /gene="repT"
                     /note="putative DnaA box required for plasmid replication"
     misc_feature    complement(3335. .3343)
                     /gene="repT"
                     /note="putative DnaA box required for plasmid replication"
     RBS             complement(4165. .4169)
                     /gene="repT"
     -10_signal      complement(4194. .4199)
                     /note="putative"
     -35_signal      complement(4218. .4223)
                     /note="putative"
     -10_signal      complement(4253. .4258)
                     /note="putative"
     -35_signal      complement(4276. .4281)
                     /note="putative"
ORIGIN

  Query Match             100.0%;  Score 5849;  DB 1;  Length 5849;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 5849;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TCTAGAAGGTCAGGGTGGACAAGGAAAACACCATAGCCCCTGCCAAGAAGATGGACGAGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TCTAGAAGGTCAGGGTGGACAAGGAAAACACCATAGCCCCTGCCAAGAAGATGGACGAGT 60

Qy         61 TGGTGTCCGGAAAAGTGGCCATCCGGGGCGCTCTTGACAACTATTTTCCAGCGGTGGCCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGGTGTCCGGAAAAGTGGCCATCCGGGGCGCTCTTGACAACTATTTTCCAGCGGTGGCCA 120

Qy        121 CCGGCATTGGCCACGAGGTACGAGCTTGTGGAGTAGACGGCCACAAAGGGGTCGTCCTCA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGCATTGGCCACGAGGTACGAGCTTGTGGAGTAGACGGCCACAAAGGGGTCGTCCTCA 180

Qy        181 AACTTCTTTTCTAGTGCCGCTTGGACGAAGGGGAGGAAGAGGAAAGGCTTCATGGCCTCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AACTTCTTTTCTAGTGCCGCTTGGACGAAGGGGAGGAAGAGGAAAGGCTTCATGGCCTCA 240

Qy        241 CCTCCTTCCCCTCCTCCTTGGCGGCCTTAGCGGCGTAAAACTCTGAGACGGCCTGAAGTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCTCCTTCCCCTCCTCCTTGGCGGCCTTAGCGGCGTAAAACTCTGAGACGGCCTGAAGTT 300

Qy        301 TAGGGATTTCGCTTTCGGGGATAAGAATCCGGCGGCTCAGGGGATGCCGGATGGCCCTTA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TAGGGATTTCGCTTTCGGGGATAAGAATCCGGCGGCTCAGGGGATGCCGGATGGCCCTTA 360

Qy        361 TCCTGCCGTCCCTTATGTACTCGTAAATGGTGGCCTTGGGTACTTTAAACCGTTCTGAAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCCTGCCGTCCCTTATGTACTCGTAAATGGTGGCCTTGGGTACTTTAAACCGTTCTGAAA 420

Qy        421 CTTCTCTAACAGAGAGCACAAAACCTCTAAAAACCTATCAATCCCACCGATTCCAGTATA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTTCTCTAACAGAGAGCACAAAACCTCTAAAAACCTATCAATCCCACCGATTCCAGTATA 480

Qy        481 CCATAAATGGCACAAAGTTTTGAGAAGGTGGTCAAACAAAAAGGCTTTCTCGGTCAGGTT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCATAAATGGCACAAAGTTTTGAGAAGGTGGTCAAACAAAAAGGCTTTCTCGGTCAGGTT 540

Qy        541 ATGGTGAGGTGGGGGCGGTCAAAGGCCGACTTAAGTTTGGTAAAGCCGGGAGGAAGCAAA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATGGTGAGGTGGGGGCGGTCAAAGGCCGACTTAAGTTTGGTAAAGCCGGGAGGAAGCAAA 600

Qy        601 CCGGGGTGTTACCATGCAACAGATGGCCGAGTGGAACGTGTGGACACAGAGAAGCGTTGA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

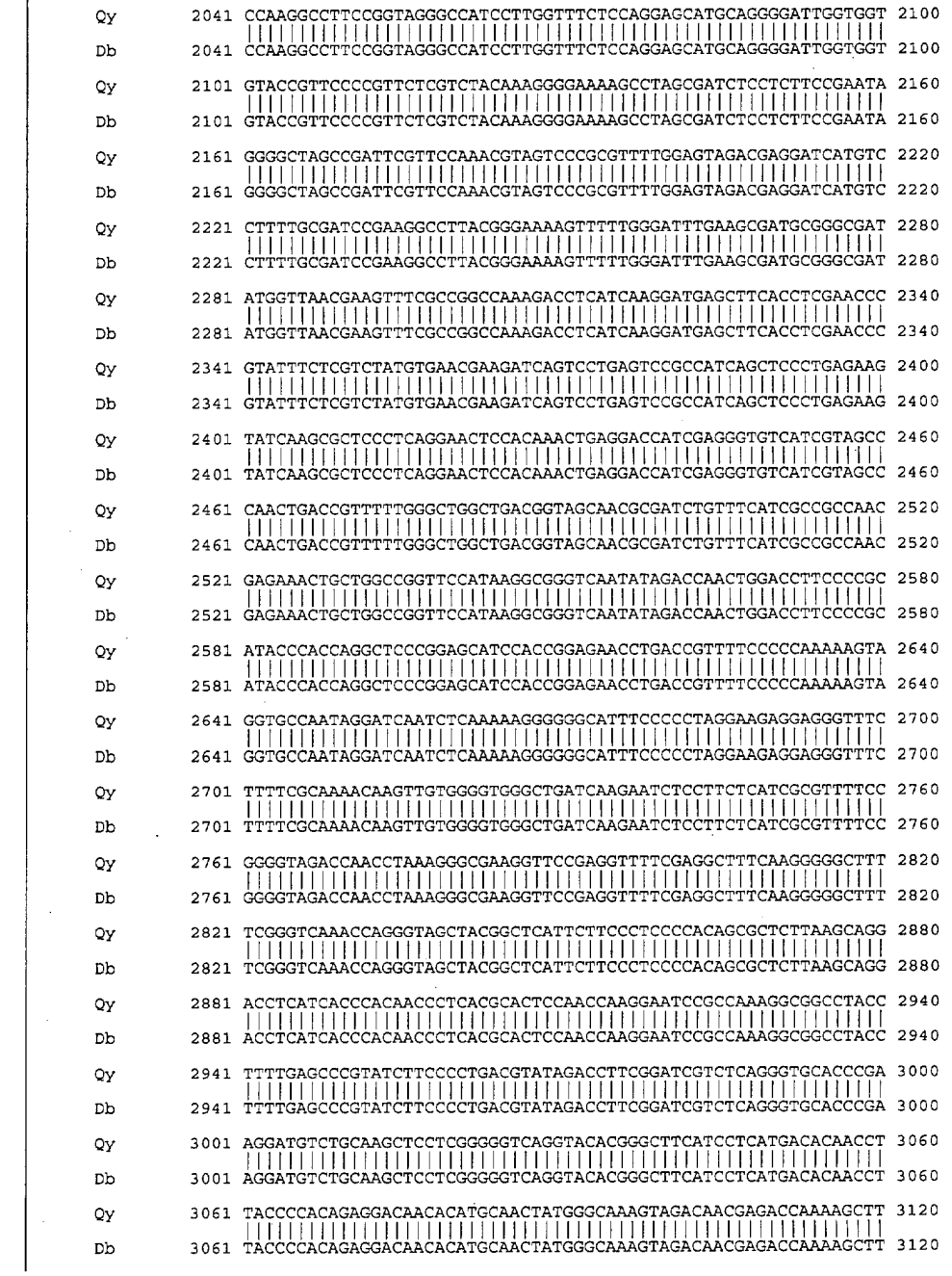

Db 601 CCGGGGTGTTACCATGCAACAGATGGCCGAGTGGAACGTGTGGACACAGAGAAGCGTTGA 660
Qy 661 GCTTCTGGAGAAGGGGTATTTGGATAAACTACTGCAGGTCTATAAAGGGGAAAGTGGCTC 720
Db 661 GCTTCTGGAGAAGGGGTATTTGGATAAACTACTGCAGGTCTATAAAGGGGAAAGTGGCTC 720
Qy 721 TTCGAGGTCAGTACCAGAGGAGGTAGAGGAAAAACTTCGCGAGGCCTACAAGGCATACGA 780
Db 721 TTCGAGGTCAGTACCAGAGGAGGTAGAGGAAAAACTTCGCGAGGCCTACAAGGCATACGA 780
Qy 781 GGGGAGGCAGGATAGTCCGGAGGCAGAAACGAAACTCGTGGAAGCCGTGCTAAATGCCAG 840
Db 781 GGGGAGGCAGGATAGTCCGGAGGCAGAAACGAAACTCGTGGAAGCCGTGCTAAATGCCAG 840
Qy 841 AAAAAAGGTCGAGCGGTCCCCCTTCAATCACCCCTACCTGCCTTTGGTCTACTACCTGGT 900
Db 841 AAAAAAGGTCGAGCGGTCCCCCTTCAATCACCCCTACCTGCCTTTGGTCTACTACCTGGT 900
Qy 901 TTCGGAAAAAGCAGAAAAAGCGAACAAGGCCCTTGAGGAGGCATTGCAGGAGGTTGCCTC 960
Db 901 TTCGGAAAAAGCAGAAAAAGCGAACAAGGCCCTTGAGGAGGCATTGCAGGAGGTTGCCTC 960
Qy 961 AAAGCACCCAGAAACCATCCGCGTCCTGGCCAAGGAAGCGCAAAGAAGAGGCGTAGAAGC 1020
Db 961 AAAGCACCCAGAAACCATCCGCGTCCTGGCCAAGGAAGCGCAAAGAAGAGGCGTAGAAGC 1020
Qy 1021 CTTGATCCAAAGGCTCAAGGAGCCTCCCGAAATAAATCGGCAGATAGGGCCGATGTTCAA 1080
Db 1021 CTTGATCCAAAGGCTCAAGGAGCCTCCCGAAATAAATCGGCAGATAGGGCCGATGTTCAA 1080
Qy 1081 AAGGTGGTACAAAGAAGAGCTAAAGGGGAAAATAGAAGAGAGGCTTCCAGGCCCTACCAA 1140
Db 1081 AAGGTGGTACAAAGAAGAGCTAAAGGGGAAAATAGAAGAGAGGCTTCCAGGCCCTACCAA 1140
Qy 1141 ACCAAAGATTGTGGTAGTATCCCCTGAAAAAAGTAAACCGGAGCAAGCACCCCTTATTGC 1200
Db 1141 ACCAAAGATTGTGGTAGTATCCCCTGAAAAAAGTAAACCGGAGCAAGCACCCCTTATTGC 1200
Qy 1201 GGAGAGAGAAGCGGGCATCATCATATACACGGGATCGGATGAAGCTTTGAAAGATGCCGC 1260
Db 1201 GGAGAGAGAAGCGGGCATCATCATATACACGGGATCGGATGAAGCTTTGAAAGATGCCGC 1260
Qy 1261 CAAGGAAAACCTGGGCCTTGGCGAGGAAGCAGAACTAGGCACCAAGGGCGTAGATTTCTA 1320
Db 1261 CAAGGAAAACCTGGGCCTTGGCGAGGAAGCAGAACTAGGCACCAAGGGCGTAGATTTCTA 1320
Qy 1321 CGTGGTCATCCGGCGTAGCCCTGAAGAGACATGGCACCTAACAGGAGAAGTGAAGTTTCA 1380
Db 1321 CGTGGTCATCCGGCGTAGCCCTGAAGAGACATGGCACCTAACAGGAGAAGTGAAGTTTCA 1380
Qy 1381 ATCCGACTTTGGCGGAAACCAAGACAACCAGAAACTAGTAGCAAAGGCTTCCATAAGGTT 1440
Db 1381 ATCCGACTTTGGCGGAAACCAAGACAACCAGAAACTAGTAGCAAAGGCTTCCATAAGGTT 1440
Qy 1441 GGACCTTGAGAAGAGGCACATAGGAATAGTGGTGGTGGACGGAATGCCTGTGGTGAGCAA 1500
Db 1441 GGACCTTGAGAAGAGGCACATAGGAATAGTGGTGGTGGACGGAATGCCTGTGGTGAGCAA 1500
Qy 1501 GTTTCGTGGGTGGGCCGGACTGGGGAAAGAAACGATCGTTACATCCGTACTCCTCCTTCC 1560
Db 1501 GTTTCGTGGGTGGGCCGGACTGGGGAAAGAAACGATCGTTACATCCGTACTCCTCCTTCC 1560
Qy 1561 AGACCTGATAGCGGAGCTCTACCAAAAGGGTGAAGAAGCCCTGGGCCTCTAGAAGGCGGA 1620
Db 1561 AGACCTGATAGCGGAGCTCTACCAAAAGGGTGAAGAAGCCCTGGGCCTCTAGAAGGCGGA 1620
Qy 1621 CACAATCTCAAACTTGTGCTGTAGCCTGGGGAAATCCTCTAACACCCTTCTAGTGAAGGC 1680
Db 1621 CACAATCTCAAACTTGTGCTGTAGCCTGGGGAAATCCTCTAACACCCTTCTAGTGAAGGC 1680
Qy 1681 TTTGACCGCCTCCCAGGAGGCATCTATGCCGATGGATCGCCGCTTTAAGAGGGGTGAGGC 1740
Db 1681 TTTGACCGCCTCCCAGGAGGCATCTATGCCGATGGATCGCCGCTTTAAGAGGGGTGAGGC 1740
Qy 1741 TATAAGCGTAGTACCGGAGCCTGCGAAGGGATCGAGCACTAAATCCCCCTCGTTACTCCC 1800
Db 1741 TATAAGCGTAGTACCGGAGCCTGCGAAGGGATCGAGCACTAAATCCCCCTCGTTACTCCC 1800
Qy 1801 TGTTTGGACGATGAGCTTGAGCATGTCCAGATTTTTCTCGGTGGGGTATCGCGGGTACGG 1860
Db 1801 TGTTTGGACGATGAGCTTGAGCATGTCCAGATTTTTCTCGGTGGGGTATCGCGGGTACGG 1860
Qy 1861 AGGATCCTTGAACTGCCAAACGTCCTGGAGCTTCTTCCCCTTCTTCAGGCGATCCCGAGC 1920
Db 1861 AGGATCCTTGAACTGCCAAACGTCCTGGAGCTTCTTCCCCTTCTTCAGGCGATCCCGAGC 1920
Qy 1921 GTAAACTTTCTTCCGCGGCACCCCGTTCTTTGACCAGACAATAAGCCCTTGAGCGTCTAG 1980
Db 1921 GTAAACTTTCTTCCGCGGCACCCCGTTCTTTGACCAGACAATAAGCCCTTGAGCGTCTAG 1980
Qy 1981 CTCGTCAAGCTTCTCCGGGGGGATAGCGCCAATGCCGTCCAGGAGGGGGAAGTATTCCTCG 2040
Db 1981 CTCGTCAAGCTTCTCCGGGGGGATAGCGCCAATGCCGTCCAGGAGGGGGAAGTATTCCTCG 2040
Qy 2041 CCAAGGCCTTCCGGTAGGGCCATCCTTGGTTTCTCCAGGAGCATGCAGGGGATTGGTGGT 2100
Db 2041 CCAAGGCCTTCCGGTAGGGCCATCCTTGGTTTCTCCAGGAGCATGCAGGGGATTGGTGGT 2100
Qy 2101 GTACCGTTCCCCGTTCTCGTCTACAAAGGGGAAAAGCCTAGCGATCTCCTCTTCCGAATA 2160
Db 2101 GTACCGTTCCCCGTTCTCGTCTACAAAGGGGAAAAGCCTAGCGATCTCCTCTTCCGAATA 2160
Qy 2161 GGGGCTAGCCGATTCGTTCCAAACGTAGTCCCGCGTTTTGGAGTAGACGAGGATCATGTC 2220
Db 2161 GGGGCTAGCCGATTCGTTCCAAACGTAGTCCCGCGTTTTGGAGTAGACGAGGATCATGTC 2220
Qy 2221 CTTTTGCGATCCGAAGGCCTTACGGGAAAAGTTTTTGGGATTTGAAGCGATGCGGGCGAT 2280
Db 2221 CTTTTGCGATCCGAAGGCCTTACGGGAAAAGTTTTTGGGATTTGAAGCGATGCGGGCGAT 2280
Qy 2281 ATGGTTAACGAAGTTTCGCCGGCCAAAGACCTCATCAAGGATGAGCTTCACCTCGAACCC 2340
Db 2281 ATGGTTAACGAAGTTTCGCCGGCCAAAGACCTCATCAAGGATGAGCTTCACCTCGAACCC 2340
Qy 2341 GTATTTCTCGTCTATGTGAACGAAGATCAGTCCTGAGTCCGCCATCAGCTCCCTGAGAAG 2400
Db 2341 GTATTTCTCGTCTATGTGAACGAAGATCAGTCCTGAGTCCGCCATCAGCTCCCTGAGAAG 2400
Qy 2401 TATCAAGCGCTCCCTCAGGAACTCCACAAACTGAGGACCATCGAGGGTGTCATCGTAGCC 2460
Db 2401 TATCAAGCGCTCCCTCAGGAACTCCACAAACTGAGGACCATCGAGGGTGTCATCGTAGCC 2460
Qy 2461 CAACTGACCGTTTTTGGGCTGGCTGACGGTAGCAACGCGATCTGTTTCATCGCCGCCAAC 2520
Db 2461 CAACTGACCGTTTTTGGGCTGGCTGACGGTAGCAACGCGATCTGTTTCATCGCCGCCAAC 2520
Qy 2521 GAGAAACTGCTGGCCGGTTCCATAAGGCGGGTCAATATAGACCAACTGGACCTTCCCCGC 2580
Db 2521 GAGAAACTGCTGGCCGGTTCCATAAGGCGGGTCAATATAGACCAACTGGACCTTCCCCGC 2580
Qy 2581 ATACCCACCAGGCTCCCGGAGCATCCACCGGAGAACCTGACCGTTTTCCCCCAAAAAGTA 2640
Db 2581 ATACCCACCAGGCTCCCGGAGCATCCACCGGAGAACCTGACCGTTTTCCCCCAAAAAGTA 2640
Qy 2641 GGTGCCAATAGGATCAATCTCAAAAAGGGGGGCATTTCCCCCTAGGAAGAGGAGGGTTTC 2700
Db 2641 GGTGCCAATAGGATCAATCTCAAAAAGGGGGGCATTTCCCCCTAGGAAGAGGAGGGTTTC 2700
Qy 2701 TTTTCGCAAAACAAGTTGTGGGGTGGGCTGATCAAGAATCTCCTTCTCATCGCGTTTTCC 2760
Db 2701 TTTTCGCAAAACAAGTTGTGGGGTGGGCTGATCAAGAATCTCCTTCTCATCGCGTTTTCC 2760
Qy 2761 GGGGTAGACCAACCTAAAGGGCGAAGGTTCCGAGGTTTTCGAGGCTTTCAAGGGGGCTTT 2820
Db 2761 GGGGTAGACCAACCTAAAGGGCGAAGGTTCCGAGGTTTTCGAGGCTTTCAAGGGGGCTTT 2820


Qy 2821 TCGGGTCAAACCAGGGTAGCTACGGCTCATTCTTCCCTCCCCACAGCGCTCTTAAGCAGG 2880
Db 2821 TCGGGTCAAACCAGGGTAGCTACGGCTCATTCTTCCCTCCCCACAGCGCTCTTAAGCAGG 2880
Qy 2881 ACCTCATCACCCACAACCCTCACGCACTCCAACCAAGGAATCCGCCAAAGGCGGCCTACC 2940
Db 2881 ACCTCATCACCCACAACCCTCACGCACTCCAACCAAGGAATCCGCCAAAGGCGGCCTACC 2940
Qy 2941 TTTTGAGCCCGTATCTTCCCCTGACGTATAGACCTTCGGATCGTCTCAGGGTGCACCCGA 3000
Db 2941 TTTTGAGCCCGTATCTTCCCCTGACGTATAGACCTTCGGATCGTCTCAGGGTGCACCCGA 3000
Qy 3001 AGGATGTCTGCAAGCTCCTCGGGGGTCAGGTACACGGGCTTCATCCTCATGACACAACCT 3060
Db 3001 AGGATGTCTGCAAGCTCCTCGGGGGTCAGGTACACGGGCTTCATCCTCATGACACAACCT 3060
Qy 3061 TACCCCACAGAGGACAACACATGCAACTATGGGCAAAGTAGACAACGAGACCAAAAGCTT 3120
Db 3061 TACCCCACAGAGGACAACACATGCAACTATGGGCAAAGTAGACAACGAGACCAAAAGCTT 3120
Qy 3121 GGGCCACTCTCTCAGGAGGCCTCCTTGAGGGTCTTCACTAGGTACGCTCCCGGTCGTGTC 3180
Db 3121 GGGCCACTCTCTCAGGAGGCCTCCTTGAGGGTCTTCACTAGGTACGCTCCCGGTCGTGTC 3180
Qy 3181 AGATGGCCATCCGTGTAATCTCGGATAACCCGCATTAGGACGGCAAATAGATATTGCGCG 3240
Db 3181 AGATGGCCATCCGTGTAATCTCGGATAACCCGCATTAGGACGGCAAATAGATATTGCGCG 3240
Qy 3241 GGGAGTTCACCCCTGGCCACAGCCCACAGCAAGCCTGCATAGAAACGTCTTGAACGACGG 3300
Db 3241 GGGAGTTCACCCCTGGCCACAGCCCACAGCAAGCCTGCATAGAAACGTCTTGAACGACGG 3300
Qy 3301 TCATCTAGGAGATCGGCAATGTACGTAGCAATGAGGGTGATAAGGGCCGGGAGTTTGGAA 3360
Db 3301 TCATCTAGGAGATCGGCAATGTACGTAGCAATGAGGGTGATAAGGGCCGGGAGTTTGGAA 3360
Qy 3361 CGCTCCACCTCGGGGAGGACCAGGATGAGGCCCAGGTCAACGGCCACGGTCTTGGTGTTG 3420
Db 3361 CGCTCCACCTCGGGGAGGACCAGGATGAGGCCCAGGTCAACGGCCACGGTCTTGGTGTTG 3420
Qy 3421 GGCATCACCCTTTTCCCCTGAGCCCAGAGGACCAGCACGTCCAGGGTGGGGCGGATTCCG 3480
Db 3421 GGCATCACCCTTTTCCCCTGAGCCCAGAGGACCAGCACGTCCAGGGTGGGGCGGATTCCG 3480
Qy 3481 TGGTCCTGGTAGGCCTTGACCCAGTTGAAGGAGAGCACGCCGTTGGCCATGTCTAGGGCG 3540
Db 3481 TGGTCCTGGTAGGCCTTGACCCAGTTGAAGGAGAGCACGCCGTTGGCCATGTCTAGGGCG 3540
Qy 3541 AGGTTCCTCCAGGGGTAGATGTAGTCGTCCAGGGTGAGCCTGGCTTTCCCTGGCCTCAGC 3600
Db 3541 AGGTTCCTCCAGGGGTAGATGTAGTCGTCCAGGGTGAGCCTGGCTTTCCCTGGCCTCAGC 3600
Qy 3601 CGGACGGCCCAAAGGGTGCCGATGGCCCGGCGCTCCCCGTTGACGGTTTGGTGAAGGACG 3660
Db 3601 CGGACGGCCCAAAGGGTGCCGATGGCCCGGCGCTCCCCGTTGACGGTTTGGTGAAGGACG 3660
Qy 3661 TCGGTGGCCACCAGGCCCTTTTTCTCAAGGACCTTCTTCCAGGCGTGGACGGTCTGCCGG 3720
Db 3661 TCGGTGGCCACCAGGCCCTTTTTCTCAAGGACCTTCTTCCAGGCGTGGACGGTCTGCCGG 3720
Qy 3721 GTGACCCCCAGGTTGAGGGCCAACATCTCCAGGGGGACCATGAAGACGACCGTCCCCACC 3780
Db 3721 GTGACCCCCAGGTTGAGGGCCAACATCTCCAGGGGGACCATGAAGACGACCGTCCCCACC 3780
Qy 3781 TTCCTCCAGAGCTCCCGGTTGCCGTAGGGGATGGTGGAGCGGGCAATCTCCTGGAGGAGT 3840
Db 3781 TTCCTCCAGAGCTCCCGGTTGCCGTAGGGGATGGTGGAGCGGGCAATCTCCTGGAGGAGT 3840
Qy 3841 TCCAGAAGCTTCTGAGGCCCGTCCTGGACGGCTTGACATACATTCCCAACGGGGGGTTCA 3900
Db 3841 TCCAGAAGCTTCTGAGGCCCGTCCTGGACGGCTTGACATACATTCCCAACGGGGGGTTCA 3900
Qy 3901 GCTGGTGGTATAGGGGGGGCCCCAACAGAGGAAAGGAAAGCCTCATGGTCTTTCTCGGGT 3960
Db 3901 GCTGGTGGTATAGGGGGGGCCCCAACAGAGGAAAGGAAAGCCTCATGGTCTTTCTCGGGT 3960
Qy 3961 AAAGCAATCCTAAGTGCCTCTTTTTGGTATGTAAAGCCCTTCGCGAGGCGATTTTCGGCA 4020
Db 3961 AAAGCAATCCTAAGTGCCTCTTTTTGGTATGTAAAGCCCTTCGCGAGGCGATTTTCGGCA 4020
Qy 4021 CCTCCATCTGGAGGGGGGTCGGTGGCCAAGAAGAAGTCCTCTGACCCCCTATCTGACCCC 4080
Db 4021 CCTCCATCTGGAGGGGGGTCGGTGGCCAAGAAGAAGTCCTCTGACCCCCTATCTGACCCC 4080
Qy 4081 CTAGTGGCATCGGTGTTGTCGTGGGTTTCCTCTAAAGCCTCGTAAAGCTCTTCAAAGAAG 4140
Db 4081 CTAGTGGCATCGGTGTTGTCGTGGGTTTCCTCTAAAGCCTCGTAAAGCTCTTCAAAGAAG 4140
Qy 4141 GTTTTTTCGTTCTTCACCCTCGGACCTCCTTGTCATCTGGAGCCCGAGGCGTTACCCTAG 4200
Db 4141 GTTTTTTCGTTCTTCACCCTCGGACCTCCTTGTCATCTGGAGCCCGAGGCGTTACCCTAG 4200
Qy 4201 GTCTTGGGGGTGATCCGGGGCAACCGCCTCGGTTTCGCCTTTTTATGGGTCCAAAATAAC 4260
Db 4201 GTCTTGGGGGTGATCCGGGGCAACCGCCTCGGTTTCGCCTTTTTATGGGTCCAAAATAAC 4260
Qy 4261 CGTCAGCCCAGCGGCTGGCAATCCCCCCTCCTAAAAGGCCGTTATAGGCCCTGCTAGGAG 4320
Db 4261 CGTCAGCCCAGCGGCTGGCAATCCCCCCTCCTAAAAGGCCGTTATAGGCCCTGCTAGGAG 4320
Qy 4321 GGGGGTAGTACTTTCCTACCCCCCTAGGCTTGGAGAGGCCTTAGGAGGTCTCCTAGGGCC 4380
Db 4321 GGGGGTAGTACTTTCCTACCCCCCTAGGCTTGGAGAGGCCTTAGGAGGTCTCCTAGGGCC 4380
Qy 4381 TCGTGGGGGTGTAGGGGTAACCTCATGGCCAGGCCGGCCGGCTCGGGACTCTGGAGGAGG 4440
Db 4381 TCGTGGGGGTGTAGGGGTAACCTCATGGCCAGGCCGGCCGGCTCGGGACTCTGGAGGAGG 4440
Qy 4441 CCTCCATAGCCTACTCGTGGTGGAGGTTTGTGAAGGGGTTCACTAATGCATACGGCTAGC 4500
Db 4441 CCTCCATAGCCTACTCGTGGTGGAGGTTTGTGAAGGGGTTCACTAATGCATACGGCTAGC 4500
Qy 4501 CTCGGGATCACGGCCAAATGGTATGCAGGTTTTGGTATAAAACCCTCAGGTTTGAGGCTA 4560
Db 4501 CTCGGGATCACGGCCAAATGGTATGCAGGTTTTGGTATAAAACCCTCAGGTTTGAGGCTA 4560
Qy 4561 GTTTATGTCGGTTTTATGCACCTTTGACTCGGATCACGGGCATAAACACCAGTTTCCTGC 4620
Db 4561 GTTTATGTCGGTTTTATGCACCTTTGACTCGGATCACGGGCATAAACACCAGTTTCCTGC 4620
Qy 4621 ACGAAAGAAAACTTTCGCGATCTAAGAGGGGGAAAGAGGTGTAGAGGGACGGCCTTCATG 4680
Db 4621 ACGAAAGAAAACTTTCGCGATCTAAGAGGGGGAAAGAGGTGTAGAGGGACGGCCTTCATG 4680
Qy 4681 AAAGTTGGCCTCTTAGGAGGCCGTTGTAGAGGGCCGTCTCGGGTTCAAATCCTTTCCCTC 4740
Db 4681 AAAGTTGGCCTCTTAGGAGGCCGTTGTAGAGGGCCGTCTCGGGTTCAAATCCTTTCCCTC 4740
Qy 4741 TCTCTCCAGGTTTCCGAGGTTCGAGGTCTTGGTCCAGGTCTTGTACCAAGTTTTTGACCA 4800
Db 4741 TCTCTCCAGGTTTCCGAGGTTCGAGGTCTTGGTCCAGGTCTTGTACCAAGTTTTTGACCA 4800
Qy 4801 AAGTCTATTCTCGGAATATAGGGGTATCTTGTCTATCTTCCCTACGGGATATCTCTGTCT 4860
Db 4801 AAGTCTATTCTCGGAATATAGGGGTATCTTGTCTATCTTCCCTACGGGATATCTCTGTCT 4860
Qy 4861 GTGTGAACTTGATCCCATCCCAATACATATCTCAATCTCCTAATCTCCTCTTCTCTCCAG 4920
Db 4861 GTGTGAACTTGATCCCATCCCAATACATATCTCAATCTCCTAATCTCCTCTTCTCTCCAG 4920
Qy 4921 ATCCCTAATCTCTTCTTCTACCTCTTTCTCCTCCCAATTAAGAATGGAGAGGAAAAACCC 4980
Db 4921 ATCCCTAATCTCTTCTTCTACCTCTTTCTCCTCCCAATTAAGAATGGAGAGGAAAAACCC 4980
Qy 4981 CGACCAGAACGAGCTTCTCGGGGTCAGTTTCGGTAATCTCGGGACAGGTTTTCATCGTCT 5040

```
Db       4981 CGACCAGAACGAGCTTCTCGGGGTCAGTTTCGGTAATCTCGGGACAGGTTTTCATCGTCT 5040

Qy       5041 AGGACGAGGATTAGGGCATGAAAAATGGGCTTTGACAAAATCTTTCTAAAAAATACTCCC 5100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5041 AGGACGAGGATTAGGGCATGAAAAATGGGCTTTGACAAAATCTTTCTAAAAAATACTCCC 5100

Qy       5101 CGAGGTTGGGGAAGTGCCCTCGGGGAGAAGATTTTTGGCAGTTTAGATGTTATGCTCTAT 5160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5101 CGAGGTTGGGGAAGTGCCCTCGGGGAGAAGATTTTTGGCAGTTTAGATGTTATGCTCTAT 5160

Qy       5161 CACGGGCCGGAGGCCTCCACGATAAGTTGTCTTGGCCAAGTACCGGGCCAGGTCGGGGGT 5220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5161 CACGGGCCGGAGGCCTCCACGATAAGTTGTCTTGGCCAAGTACCGGGCCAGGTCGGGGGT 5220

Qy       5221 GCTCTTCAGCGTGGTGATGGTACTTTCACGGAAGTTCACAAGTCCTTTTAGAGGCTTCAG 5280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5221 GCTCTTCAGCGTGGTGATGGTACTTTCACGGAAGTTCACAAGTCCTTTTAGAGGCTTCAG 5280

Qy       5281 GTCGGGGATAGTGCTCAAGTACTCCCAAGCGTTCTCGGGCCCGTGGTCGGGGAGAAGGAC 5340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5281 GTCGGGGATAGTGCTCAAGTACTCCCAAGCGTTCTCGGGCCCGTGGTCGGGGAGAAGGAC 5340

Qy       5341 AAAGGGGTCGGGCAAAAGTTCATCTTTGTACTTAGGACGGATTACTTTAGCACCTGATAA 5400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5341 AAAGGGGTCGGGCAAAAGTTCATCTTTGTACTTAGGACGGATTACTTTAGCACCTGATAA 5400

Qy       5401 CTTCAGGGCCGTTAAGAAGGGCCTCACCTCGGAGACGGGTGGAAGGAGGACGTGGGCGTG 5460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5401 CTTCAGGGCCGTTAAGAAGGGCCTCACCTCGGAGACGGGTGGAAGGAGGACGTGGGCGTG 5460

Qy       5461 GAAGAAGACGAACCCCGATTTTTGGGAAGTCTCCCTCCAGTTTGATGATGAACGTTGGGA 5520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5461 GAAGAAGACGAACCCCGATTTTTGGGAAGTCTCCCTCCAGTTTGATGATGAACGTTGGGA 5520

Qy       5521 GGAAGCCGGCCAGGATGTCTTTCATCGCGCCTCGAACCTCGGACACATAAAAAACTTTCG 5580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5521 GGAAGCCGGCCAGGATGTCTTTCATCGCGCCTCGAACCTCGGACACATAAAAAACTTTCG 5580

Qy       5581 TGTTTGTCAGGGCAAGAGTGCTATGTATGAGGTAACCTTCGGGAGTACAAAGTGCCTCAA 5640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5581 TGTTTGTCAGGGCAAGAGTGCTATGTATGAGGTAACCTTCGGGAGTACAAAGTGCCTCAA 5640

Qy       5641 GCCGCCTTTCCCAACGCTCCAAAACTCTAGGGTCAGGTGGTTTAGGTTTTCTGAAAAACT 5700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5641 GCCGCCTTTCCCAACGCTCCAAAACTCTAGGGTCAGGTGGTTTAGGTTTTCTGAAAAACT 5700

Qy       5701 CTAGCTTTTCAGTGGTCATTCCTCACCCCTCTAGCACGTACTCTGGAAGGTAAACCTTTG 5760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5701 CTAGCTTTTCAGTGGTCATTCCTCACCCCTCTAGCACGTACTCTGGAAGGTAAACCTTTG 5760

Qy       5761 ACACAGCGGCCAAGTCTAGCGTCTCCCAGTCCAGTTGGTCTGGGACGCGTGAGAAGGGGA 5820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5761 ACACAGCGGCCAAGTCTAGCGTCTCCCAGTCCAGTTGGTCTGGGACGCGTGAGAAGGGGA 5820

Qy       5821 GGGGCTTGGTGTAGAGGACCAGAAGACCC 5849
             |||||||||||||||||||||||||||||
Db       5821 GGGGCTTGGTGTAGAGGACCAGAAGACCC 5849
```

```
RESULT 2
AR139426
LOCUS       AR139426               5849 bp    DNA     linear   PAT 16-JUN-2001
DEFINITION  Sequence 6 from patent US 6207377.
ACCESSION   AR139426
VERSION     AR139426.1  GI:14481922
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 5849)
  AUTHORS   Wayne,J. and Xu,S.-y.
  TITLE     Method for construction of thermus-E. coli shuttle vectors and
            identification of two Thermus plasmid replication origins
  JOURNAL   Patent: US 6207377-A 6 27-MAR-2001;
FEATURES             Location/Qualifiers
     source          1. .5849
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match           100.0%;  Score 5849;  DB 6;  Length 5849;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 5849;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TCTAGAAGGTCAGGGTGGACAAGGAAAACACCATAGCCCCTGCCAAGAAGATGGACGAGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TCTAGAAGGTCAGGGTGGACAAGGAAAACACCATAGCCCCTGCCAAGAAGATGGACGAGT 60

Qy         61 TGGTGTCCGGAAAAGTGGCCATCCGGGGCGCTCTTGACAACTATTTTCCAGCGGTGGCCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGGTGTCCGGAAAAGTGGCCATCCGGGGCGCTCTTGACAACTATTTTCCAGCGGTGGCCA 120

Qy        121 CCGGCATTGGCCACGAGGTACGAGCTTGTGGAGTAGACGGCCACAAAGGGGTCGTCCTCA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGCATTGGCCACGAGGTACGAGCTTGTGGAGTAGACGGCCACAAAGGGGTCGTCCTCA 180

Qy        181 AACTTCTTTTCTAGTGCCGCTTGGACGAAGGGGAGGAAGAGGAAAGGCTTCATGGCCTCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AACTTCTTTTCTAGTGCCGCTTGGACGAAGGGGAGGAAGAGGAAAGGCTTCATGGCCTCA 240

Qy        241 CCTCCTTCCCCTCCTCCTTGGCGGCCTTAGCGGCGTAAAACTCTGAGACGGCCTGAAGTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCTCCTTCCCCTCCTCCTTGGCGGCCTTAGCGGCGTAAAACTCTGAGACGGCCTGAAGTT 300

Qy        301 TAGGGATTTCGCTTTCGGGGATAAGAATCCGGCGGCTCAGGGGATGCCGGATGGCCCTTA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TAGGGATTTCGCTTTCGGGGATAAGAATCCGGCGGCTCAGGGGATGCCGGATGGCCCTTA 360

Qy        361 TCCTGCCGTCCCTTATGTACTCGTAAATGGTGGCCTTGGGTACTTTAAACCGTTCTGAAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCCTGCCGTCCCTTATGTACTCGTAAATGGTGGCCTTGGGTACTTTAAACCGTTCTGAAA 420

Qy        421 CTTCTCTAACAGAGAGCACAAAACCTCTAAAAACCTATCAATCCCACCGATTCCAGTATA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTTCTCTAACAGAGAGCACAAAACCTCTAAAAACCTATCAATCCCACCGATTCCAGTATA 480

Qy        481 CCATAAATGGCACAAAGTTTTGAGAAGGTGGTCAAACAAAAAGGCTTTCTCGGTCAGGTT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCATAAATGGCACAAAGTTTTGAGAAGGTGGTCAAACAAAAAGGCTTTCTCGGTCAGGTT 540

Qy        541 ATGGTGAGGTGGGGGCGGTCAAAGGCCGACTTAAGTTTGGTAAAGCCGGGAGGAAGCAAA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATGGTGAGGTGGGGGCGGTCAAAGGCCGACTTAAGTTTGGTAAAGCCGGGAGGAAGCAAA 600

Qy        601 CCGGGGTGTTACCATGCAACAGATGGCCGAGTGGAACGTGTGGACACAGAGAAGCGTTGA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCGGGGTGTTACCATGCAACAGATGGCCGAGTGGAACGTGTGGACACAGAGAAGCGTTGA 660

Qy        661 GCTTCTGGAGAAGGGGTATTTGGATAAACTACTGCAGGTCTATAAAGGGGAAAGTGGCTC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GCTTCTGGAGAAGGGGTATTTGGATAAACTACTGCAGGTCTATAAAGGGGAAAGTGGCTC 720

Qy        721 TTCGAGGTCAGTACCAGAGGAGGTAGAGGAAAAACTTCGCGAGGCCTACAAGGCATACGA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TTCGAGGTCAGTACCAGAGGAGGTAGAGGAAAAACTTCGCGAGGCCTACAAGGCATACGA 780

Qy        781 GGGGAGGCAGGATAGTCCGGAGGCAGAAACGAAACTCGTGGAAGCCGTGCTAAATGCCAG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GGGGAGGCAGGATAGTCCGGAGGCAGAAACGAAACTCGTGGAAGCCGTGCTAAATGCCAG 840

Qy        841 AAAAAAGGTCGAGCGGTCCCCCTTCAATCACCCCTACCTGCCTTTGGTCTACTACCTGGT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAAAAAGGTCGAGCGGTCCCCCTTCAATCACCCCTACCTGCCTTTGGTCTACTACCTGGT 900

Qy        901 TTCGGAAAAAGCAGAAAAAGCGAACAAGGCCCTTGAGGAGGCATTGCAGGAGGTTGCCTC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```


Db 901 TTCGGAAAAAGCAGAAAAAGCGAACAAGGCCCTTGAGGAGGCATTGCAGGAGGTTGCCTC 960
Qy 961 AAAGCACCCAGAAACCATCCGCGTCCTGGCCAAGGAAGCGCAAAGAAGAGGCGTAGAAGC 1020
Db 961 AAAGCACCCAGAAACCATCCGCGTCCTGGCCAAGGAAGCGCAAAGAAGAGGCGTAGAAGC 1020
Qy 1021 CTTGATCCAAAGGCTCAAGGAGCCTCCCGAAATAAATCGGCAGATAGGGCCGATGTTCAA 1080
Db 1021 CTTGATCCAAAGGCTCAAGGAGCCTCCCGAAATAAATCGGCAGATAGGGCCGATGTTCAA 1080
Qy 1081 AAGGTGGTACAAAGAAGAGCTAAAGGGGAAAATAGAAGAGAGGCTTCCAGGCCCTACCAA 1140
Db 1081 AAGGTGGTACAAAGAAGAGCTAAAGGGGAAAATAGAAGAGAGGCTTCCAGGCCCTACCAA 1140
Qy 1141 ACCAAAGATTGTGGTAGTATCCCCTGAAAAAAGTAAACCGGAGCAAGCACCCCTTATTGC 1200
Db 1141 ACCAAAGATTGTGGTAGTATCCCCTGAAAAAAGTAAACCGGAGCAAGCACCCCTTATTGC 1200
Qy 1201 GGAGAGAGAAGCGGGCATCATCATATACACGGGATCGGATGAAGCTTTGAAAGATGCCGC 1260
Db 1201 GGAGAGAGAAGCGGGCATCATCATATACACGGGATCGGATGAAGCTTTGAAAGATGCCGC 1260
Qy 1261 CAAGGAAAACCTGGGCCTTGGCGAGGAAGCAGAACTAGGCACCAAGGGCGTAGATTTCTA 1320
Db 1261 CAAGGAAAACCTGGGCCTTGGCGAGGAAGCAGAACTAGGCACCAAGGGCGTAGATTTCTA 1320
Qy 1321 CGTGGTCATCCGGCGTAGCCCTGAAGAGACATGGCACCTAACAGGAGAAGTGAAGTTTCA 1380
Db 1321 CGTGGTCATCCGGCGTAGCCCTGAAGAGACATGGCACCTAACAGGAGAAGTGAAGTTTCA 1380
Qy 1381 ATCCGACTTTGGCGGAAACCAAGACAACCAGAAACTAGTAGCAAAGGCTTCCATAAGGTT 1440
Db 1381 ATCCGACTTTGGCGGAAACCAAGACAACCAGAAACTAGTAGCAAAGGCTTCCATAAGGTT 1440
Qy 1441 GGACCTTGAGAAGAGGCACATAGGAATAGTGGTGGTGGACGGAATGCCTGTGGTGAGCAA 1500
Db 1441 GGACCTTGAGAAGAGGCACATAGGAATAGTGGTGGTGGACGGAATGCCTGTGGTGAGCAA 1500
Qy 1501 GTTTCGTGGGTGGGCCGGACTGGGGAAAGAAACGATCGTTACATCCGTACTCCTCCTTCC 1560
Db 1501 GTTTCGTGGGTGGGCCGGACTGGGGAAAGAAACGATCGTTACATCCGTACTCCTCCTTCC 1560
Qy 1561 AGACCTGATAGCGGAGCTCTACCAAAAGGGTGAAGAAGCCCTGGGCCTCTAGAAGGCGGA 1620
Db 1561 AGACCTGATAGCGGAGCTCTACCAAAAGGGTGAAGAAGCCCTGGGCCTCTAGAAGGCGGA 1620
Qy 1621 CACAATCTCAAACTTGTGCTGTAGCCTGGGGAAATCCTCTAACACCCTTCTAGTGAAGGC 1680
Db 1621 CACAATCTCAAACTTGTGCTGTAGCCTGGGGAAATCCTCTAACACCCTTCTAGTGAAGGC 1680
Qy 1681 TTTGACCGCCTCCCAGGAGGCATCTATGCCGATGGATCGCCGCTTTAAGAGGGGTGAGGC 1740
Db 1681 TTTGACCGCCTCCCAGGAGGCATCTATGCCGATGGATCGCCGCTTTAAGAGGGGTGAGGC 1740
Qy 1741 TATAAGCGTAGTACCGGAGCCTGCGAAGGGATCGAGCACTAAATCCCCCTCGTTACTCCC 1800
Db 1741 TATAAGCGTAGTACCGGAGCCTGCGAAGGGATCGAGCACTAAATCCCCCTCGTTACTCCC 1800
Qy 1801 TGTTTGGACGATGAGCTTGAGCATGTCCAGATTTTTCTCGGTGGGGTATCGCGGGTACGG 1860
Db 1801 TGTTTGGACGATGAGCTTGAGCATGTCCAGATTTTTCTCGGTGGGGTATCGCGGGTACGG 1860
Qy 1861 AGGATCCTTGAACTGCCAAACGTCCTGGAGCTTCTTCCCCTTCTTCAGGCGATCCCGAGC 1920
Db 1861 AGGATCCTTGAACTGCCAAACGTCCTGGAGCTTCTTCCCCTTCTTCAGGCGATCCCGAGC 1920
Qy 1921 GTAAACTTTCTTCCGCGGCACCCCGTTCTTTGACCAGACAATAAGCCCTTGAGCGTCTAG 1980
Db 1921 GTAAACTTTCTTCCGCGGCACCCCGTTCTTTGACCAGACAATAAGCCCTTGAGCGTCTAG 1980
Qy 1981 CTCGTCAAGCTTCTCCGGGGGGATAGCGCCAATGCCGTCCAGGAGGGGGAAGTATTCCTCG 2040
Db 1981 CTCGTCAAGCTTCTCCGGGGGGATAGCGCCAATGCCGTCCAGGAGGGGGAAGTATTCCTCG 2040


Qy 2041 CCAAGGCCTTCCGGTAGGGCCATCCTTGGTTTCTCCAGGAGCATGCAGGGGATTGGTGGT 2100
Db 2041 CCAAGGCCTTCCGGTAGGGCCATCCTTGGTTTCTCCAGGAGCATGCAGGGGATTGGTGGT 2100
Qy 2101 GTACCGTTCCCCGTTCTCGTCTACAAAGGGGAAAAGCCTAGCGATCTCCTCTTCCGAATA 2160
Db 2101 GTACCGTTCCCCGTTCTCGTCTACAAAGGGGAAAAGCCTAGCGATCTCCTCTTCCGAATA 2160
Qy 2161 GGGGCTAGCCGATTCGTTCCAAACGTAGTCCCGCGTTTTGGAGTAGACGAGGATCATGTC 2220
Db 2161 GGGGCTAGCCGATTCGTTCCAAACGTAGTCCCGCGTTTTGGAGTAGACGAGGATCATGTC 2220
Qy 2221 CTTTTGCGATCCGAAGGCCTTACGGGAAAAGTTTTTGGGATTTGAAGCGATGCGGGCGAT 2280
Db 2221 CTTTTGCGATCCGAAGGCCTTACGGGAAAAGTTTTTGGGATTTGAAGCGATGCGGGCGAT 2280
Qy 2281 ATGGTTAACGAAGTTTCGCCGGCCAAAGACCTCATCAAGGATGAGCTTCACCTCGAACCC 2340
Db 2281 ATGGTTAACGAAGTTTCGCCGGCCAAAGACCTCATCAAGGATGAGCTTCACCTCGAACCC 2340
Qy 2341 GTATTTCTCGTCTATGTGAACGAAGATCAGTCCTGAGTCCGCCATCAGCTCCCTGAGAAG 2400
Db 2341 GTATTTCTCGTCTATGTGAACGAAGATCAGTCCTGAGTCCGCCATCAGCTCCCTGAGAAG 2400
Qy 2401 TATCAAGCGCTCCCTCAGGAACTCCACAAACTGAGGACCATCGAGGGTGTCATCGTAGCC 2460
Db 2401 TATCAAGCGCTCCCTCAGGAACTCCACAAACTGAGGACCATCGAGGGTGTCATCGTAGCC 2460
Qy 2461 CAACTGACCGTTTTTGGGCTGGCTGACGGTAGCAACGCGATCTGTTTCATCGCCGCCAAC 2520
Db 2461 CAACTGACCGTTTTTGGGCTGGCTGACGGTAGCAACGCGATCTGTTTCATCGCCGCCAAC 2520
Qy 2521 GAGAAACTGCTGGCCGGTTCCATAAGGCGGGTCAATATAGACCAACTGGACCTTCCCCGC 2580
Db 2521 GAGAAACTGCTGGCCGGTTCCATAAGGCGGGTCAATATAGACCAACTGGACCTTCCCCGC 2580
Qy 2581 ATACCCACCAGGCTCCCGGAGCATCCACCGGAGAACCTGACCGTTTTCCCCCAAAAAGTA 2640
Db 2581 ATACCCACCAGGCTCCCGGAGCATCCACCGGAGAACCTGACCGTTTTCCCCCAAAAAGTA 2640
Qy 2641 GGTGCCAATAGGATCAATCTCAAAAAGGGGGGCATTTCCCCCTAGGAAGAGGAGGGTTTC 2700
Db 2641 GGTGCCAATAGGATCAATCTCAAAAAGGGGGGCATTTCCCCCTAGGAAGAGGAGGGTTTC 2700
Qy 2701 TTTTCGCAAAACAAGTTGTGGGGTGGGCTGATCAAGAATCTCCTTCTCATCGCGTTTTCC 2760
Db 2701 TTTTCGCAAAACAAGTTGTGGGGTGGGCTGATCAAGAATCTCCTTCTCATCGCGTTTTCC 2760
Qy 2761 GGGGTAGACCAACCTAAAGGGCGAAGGTTCCGAGGTTTTCGAGGCTTTCAAGGGGGCTTT 2820
Db 2761 GGGGTAGACCAACCTAAAGGGCGAAGGTTCCGAGGTTTTCGAGGCTTTCAAGGGGGCTTT 2820
Qy 2821 TCGGGTCAAACCAGGGTAGCTACGGCTCATTCTTCCCTCCCCACAGCGCTCTTAAGCAGG 2880
Db 2821 TCGGGTCAAACCAGGGTAGCTACGGCTCATTCTTCCCTCCCCACAGCGCTCTTAAGCAGG 2880
Qy 2881 ACCTCATCACCCACAACCCTCACGCACTCCAACCAAGGAATCCGCCAAAGGCGGCCTACC 2940
Db 2881 ACCTCATCACCCACAACCCTCACGCACTCCAACCAAGGAATCCGCCAAAGGCGGCCTACC 2940
Qy 2941 TTTTGAGCCCGTATCTTCCCCTGACGTATAGACCTTCGGATCGTCTCAGGGTGCACCCGA 3000
Db 2941 TTTTGAGCCCGTATCTTCCCCTGACGTATAGACCTTCGGATCGTCTCAGGGTGCACCCGA 3000
Qy 3001 AGGATGTCTGCAAGCTCCTCGGGGGTCAGGTACACGGGCTTCATCCTCATGACACAACCT 3060
Db 3001 AGGATGTCTGCAAGCTCCTCGGGGGTCAGGTACACGGGCTTCATCCTCATGACACAACCT 3060
Qy 3061 TACCCCACAGAGGACAACACATGCAACTATGGGCAAAGTAGACAACGAGACCAAAAGCTT 3120
Db 3061 TACCCCACAGAGGACAACACATGCAACTATGGGCAAAGTAGACAACGAGACCAAAAGCTT 3120

```
Qy       3121 GGGCCACTCTCTCAGGAGGCCTCCTTGAGGGTCTTCACTAGGTACGCTCCCGGTCGTGTC 3180
Db       3121 GGGCCACTCTCTCAGGAGGCCTCCTTGAGGGTCTTCACTAGGTACGCTCCCGGTCGTGTC 3180

Qy       3181 AGATGGCCATCCGTGTAATCTCGGATAACCCGCATTAGGACGGCAAATAGATATTGCGCG 3240
Db       3181 AGATGGCCATCCGTGTAATCTCGGATAACCCGCATTAGGACGGCAAATAGATATTGCGCG 3240

Qy       3241 GGGAGTTCACCCCTGGCCACAGCCCACAGCAAGCCTGCATAGAAACGTCTTGAACGACGG 3300
Db       3241 GGGAGTTCACCCCTGGCCACAGCCCACAGCAAGCCTGCATAGAAACGTCTTGAACGACGG 3300

Qy       3301 TCATCTAGGAGATCGGCAATGTACGTAGCAATGAGGGTGATAAGGGCCGGGAGTTTGGAA 3360
Db       3301 TCATCTAGGAGATCGGCAATGTACGTAGCAATGAGGGTGATAAGGGCCGGGAGTTTGGAA 3360

Qy       3361 CGCTCCACCTCGGGGAGGACCAGGATGAGGCCCAGGTCAACGGCCACGGTCTTGGTGTTG 3420
Db       3361 CGCTCCACCTCGGGGAGGACCAGGATGAGGCCCAGGTCAACGGCCACGGTCTTGGTGTTG 3420

Qy       3421 GGCATCACCCTTTTCCCCTGAGCCCAGAGGACCAGCACGTCCAGGGTGGGGCGGATTCCG 3480
Db       3421 GGCATCACCCTTTTCCCCTGAGCCCAGAGGACCAGCACGTCCAGGGTGGGGCGGATTCCG 3480

Qy       3481 TGGTCCTGGTAGGCCTTGACCCAGTTGAAGGAGAGCACGCCGTTGGCCATGTCTAGGGCG 3540
Db       3481 TGGTCCTGGTAGGCCTTGACCCAGTTGAAGGAGAGCACGCCGTTGGCCATGTCTAGGGCG 3540

Qy       3541 AGGTTCCTCCAGGGGTAGATGTAGTCGTCCAGGGTGAGCCTGGCTTTCCCTGGCCTCAGC 3600
Db       3541 AGGTTCCTCCAGGGGTAGATGTAGTCGTCCAGGGTGAGCCTGGCTTTCCCTGGCCTCAGC 3600

Qy       3601 CGGACGGCCCAAAGGGTGCCGATGGCCCGGCGCTCCCCGTTGACGGTTTGGTGAAGGACG 3660
Db       3601 CGGACGGCCCAAAGGGTGCCGATGGCCCGGCGCTCCCCGTTGACGGTTTGGTGAAGGACG 3660

Qy       3661 TCGGTGGCCACCAGGCCCTTTTTCTCAAGGACCTTCTTCCAGGCGTGGACGGTCTGCCGG 3720
Db       3661 TCGGTGGCCACCAGGCCCTTTTTCTCAAGGACCTTCTTCCAGGCGTGGACGGTCTGCCGG 3720

Qy       3721 GTGACCCCCAGGTTGAGGGCCAACATCTCCAGGGGGACCATGAAGACGACCGTCCCCACC 3780
Db       3721 GTGACCCCCAGGTTGAGGGCCAACATCTCCAGGGGGACCATGAAGACGACCGTCCCCACC 3780

Qy       3781 TTCCTCCAGAGCTCCCGGTTGCCGTAGGGGATGGTGGAGCGGGCAATCTCCTGGAGGAGT 3840
Db       3781 TTCCTCCAGAGCTCCCGGTTGCCGTAGGGGATGGTGGAGCGGGCAATCTCCTGGAGGAGT 3840

Qy       3841 TCCAGAAGCTTCTGAGGCCCGTCCTGGACGGCTTGACATACATTCCCAACGGGGGGTTCA 3900
Db       3841 TCCAGAAGCTTCTGAGGCCCGTCCTGGACGGCTTGACATACATTCCCAACGGGGGGTTCA 3900

Qy       3901 GCTGGTGGTATAGGGGGGGCCCCAACAGAGGAAAGGAAAGCCTCATGGTCTTTCTCGGGT 3960
Db       3901 GCTGGTGGTATAGGGGGGGCCCCAACAGAGGAAAGGAAAGCCTCATGGTCTTTCTCGGGT 3960

Qy       3961 AAAGCAATCCTAAGTGCCTCTTTTTGGTATGTAAAGCCCTTCGCGAGGCGATTTTCGGCA 4020
Db       3961 AAAGCAATCCTAAGTGCCTCTTTTTGGTATGTAAAGCCCTTCGCGAGGCGATTTTCGGCA 4020

Qy       4021 CCTCCATCTGGAGGGGGGTCGGTGGCCAAGAAGAAGTCCTCTGACCCCCTATCTGACCCC 4080
Db       4021 CCTCCATCTGGAGGGGGGTCGGTGGCCAAGAAGAAGTCCTCTGACCCCCTATCTGACCCC 4080

Qy       4081 CTAGTGGCATCGGTGTTGTCGTGGGTTTCCTCTAAAGCCTCGTAAAGCTCTTCAAAGAAG 4140
Db       4081 CTAGTGGCATCGGTGTTGTCGTGGGTTTCCTCTAAAGCCTCGTAAAGCTCTTCAAAGAAG 4140

Qy       4141 GTTTTTTCGTTCTTCACCCTCGGACCTCCTTGTCATCTGGAGCCCGAGGCGTTACCCTAG 4200
Db       4141 GTTTTTTCGTTCTTCACCCTCGGACCTCCTTGTCATCTGGAGCCCGAGGCGTTACCCTAG 4200

Qy       4201 GTCTTGGGGGTGATCCGGGGCAACCGCCTCGGTTTCGCCTTTTTATGGGTCCAAAATAAC 4260
Db       4201 GTCTTGGGGGTGATCCGGGGCAACCGCCTCGGTTTCGCCTTTTTATGGGTCCAAAATAAC 4260

Qy       4261 CGTCAGCCCAGCGGCTGGCAATCCCCCCTCCTAAAAGGCCGTTATAGGCCCTGCTAGGAG 4320
Db       4261 CGTCAGCCCAGCGGCTGGCAATCCCCCCTCCTAAAAGGCCGTTATAGGCCCTGCTAGGAG 4320

Qy       4321 GGGGGTAGTACTTTCCTACCCCCCTAGGCTTGGAGAGGCCTTAGGAGGTCTCCTAGGGCC 4380
Db       4321 GGGGGTAGTACTTTCCTACCCCCCTAGGCTTGGAGAGGCCTTAGGAGGTCTCCTAGGGCC 4380

Qy       4381 TCGTGGGGGTGTAGGGGTAACCTCATGGCCAGGCCGGCCGGCTCGGGACTCTGGAGGAGG 4440
Db       4381 TCGTGGGGGTGTAGGGGTAACCTCATGGCCAGGCCGGCCGGCTCGGGACTCTGGAGGAGG 4440

Qy       4441 CCTCCATAGCCTACTCGTGGTGGAGGTTTGTGAAGGGGTTCACTAATGCATACGGCTAGC 4500
Db       4441 CCTCCATAGCCTACTCGTGGTGGAGGTTTGTGAAGGGGTTCACTAATGCATACGGCTAGC 4500

Qy       4501 CTCGGGATCACGGCCAAATGGTATGCAGGTTTTGGTATAAAACCCTCAGGTTTGAGGCTA 4560
Db       4501 CTCGGGATCACGGCCAAATGGTATGCAGGTTTTGGTATAAAACCCTCAGGTTTGAGGCTA 4560

Qy       4561 GTTTATGTCGGTTTTATGCACCTTTGACTCGGATCACGGGCATAAACACCAGTTTCCTGC 4620
Db       4561 GTTTATGTCGGTTTTATGCACCTTTGACTCGGATCACGGGCATAAACACCAGTTTCCTGC 4620

Qy       4621 ACGAAAGAAAACTTTCGCGATCTAAGAGGGGGAAAGAGGTGTAGAGGGACGGCCTTCATG 4680
Db       4621 ACGAAAGAAAACTTTCGCGATCTAAGAGGGGGAAAGAGGTGTAGAGGGACGGCCTTCATG 4680

Qy       4681 AAAGTTGGCCTCTTAGGAGGCCGTTGTAGAGGGCCGTCTCGGGTTCAAATCCTTTCCCTC 4740
Db       4681 AAAGTTGGCCTCTTAGGAGGCCGTTGTAGAGGGCCGTCTCGGGTTCAAATCCTTTCCCTC 4740

Qy       4741 TCTCTCCAGGTTTCCGAGGTTCGAGGTCTTGGTCCAGGTCTTGTACCAAGTTTTTGACCA 4800
Db       4741 TCTCTCCAGGTTTCCGAGGTTCGAGGTCTTGGTCCAGGTCTTGTACCAAGTTTTTGACCA 4800

Qy       4801 AAGTCTATTCTCGGAATATAGGGGTATCTTGTCTATCTTCCCTACGGGATATCTCTGTCT 4860
Db       4801 AAGTCTATTCTCGGAATATAGGGGTATCTTGTCTATCTTCCCTACGGGATATCTCTGTCT 4860

Qy       4861 GTGTGAACTTGATCCCATCCCAATACATATCTCAATCTCCTAATCTCCTCTTCTCTCCAG 4920
Db       4861 GTGTGAACTTGATCCCATCCCAATACATATCTCAATCTCCTAATCTCCTCTTCTCTCCAG 4920

Qy       4921 ATCCCTAATCTCTTCTTCTACCTCTTTCTCCTCCCAATTAAGAATGGAGAGGAAAAACCC 4980
Db       4921 ATCCCTAATCTCTTCTTCTACCTCTTTCTCCTCCCAATTAAGAATGGAGAGGAAAAACCC 4980

Qy       4981 CGACCAGAACGAGCTTCTCGGGGTCAGTTTCGGTAATCTCGGGACAGGTTTTCATCGTCT 5040
Db       4981 CGACCAGAACGAGCTTCTCGGGGTCAGTTTCGGTAATCTCGGGACAGGTTTTCATCGTCT 5040

Qy       5041 AGGACGAGGATTAGGGCATGAAAAATGGGCTTTGACAAAATCTTTCTAAAAAATACTCCC 5100
Db       5041 AGGACGAGGATTAGGGCATGAAAAATGGGCTTTGACAAAATCTTTCTAAAAAATACTCCC 5100

Qy       5101 CGAGGTTGGGGAAGTGCCCTCGGGGAGAAGATTTTTGGCAGTTTAGATGTTATGCTCTAT 5160
Db       5101 CGAGGTTGGGGAAGTGCCCTCGGGGAGAAGATTTTTGGCAGTTTAGATGTTATGCTCTAT 5160

Qy       5161 CACGGGCCGGAGGCCTCCACGATAAGTTGTCTTGGCCAAGTACCGGGCCAGGTCGGGGGT 5220
Db       5161 CACGGGCCGGAGGCCTCCACGATAAGTTGTCTTGGCCAAGTACCGGGCCAGGTCGGGGGT 5220

Qy       5221 GCTCTTCAGCGTGGTGATGGTACTTTCACGGAAGTTCACAAGTCCTTTTAGAGGCTTCAG 5280
Db       5221 GCTCTTCAGCGTGGTGATGGTACTTTCACGGAAGTTCACAAGTCCTTTTAGAGGCTTCAG 5280

Qy       5281 GTCGGGGATAGTGCTCAAGTACTCCCAAGCGTTCTCGGGCCCGTGGTCGGGGAGAAGGAC 5340
```

```
Db       5281 GTCGGGGATAGTGCTCAAGTACTCCCAAGCGTTCTCGGGCCCGTGGTCGGGGAGAAGGAC 5340

Qy       5341 AAAGGGGTCGGGCAAAAGTTCATCTTTGTACTTAGGACGGATTACTTTAGCACCTGATAA 5400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5341 AAAGGGGTCGGGCAAAAGTTCATCTTTGTACTTAGGACGGATTACTTTAGCACCTGATAA 5400

Qy       5401 CTTCAGGGCCGTTAAGAAGGGCCTCACCTCGGAGACGGGTGGAAGGAGGACGTGGGCGTG 5460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5401 CTTCAGGGCCGTTAAGAAGGGCCTCACCTCGGAGACGGGTGGAAGGAGGACGTGGGCGTG 5460

Qy       5461 GAAGAAGACGAACCCCGATTTTTGGGAAGTCTCCCTCCAGTTTGATGATGAACGTTGGGA 5520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5461 GAAGAAGACGAACCCCGATTTTTGGGAAGTCTCCCTCCAGTTTGATGATGAACGTTGGGA 5520

Qy       5521 GGAAGCCGGCCAGGATGTCTTTCATCGCGCCTCGAACCTCGGACACATAAAAAACTTTCG 5580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5521 GGAAGCCGGCCAGGATGTCTTTCATCGCGCCTCGAACCTCGGACACATAAAAAACTTTCG 5580

Qy       5581 TGTTTGTCAGGGCAAGAGTGCTATGTATGAGGTAACCTTCGGGAGTACAAAGTGCCTCAA 5640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5581 TGTTTGTCAGGGCAAGAGTGCTATGTATGAGGTAACCTTCGGGAGTACAAAGTGCCTCAA 5640

Qy       5641 GCCGCCTTTCCCAACGCTCCAAAACTCTAGGGTCAGGTGGTTTAGGTTTTCTGAAAAACT 5700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5641 GCCGCCTTTCCCAACGCTCCAAAACTCTAGGGTCAGGTGGTTTAGGTTTTCTGAAAAACT 5700

Qy       5701 CTAGCTTTTCAGTGGTCATTCCTCACCCCTCTAGCACGTACTCTGGAAGGTAAACCTTTG 5760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5701 CTAGCTTTTCAGTGGTCATTCCTCACCCCTCTAGCACGTACTCTGGAAGGTAAACCTTTG 5760

Qy       5761 ACACAGCGGCCAAGTCTAGCGTCTCCCAGTCCAGTTGGTCTGGGACGCGTGAGAAGGGGA 5820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5761 ACACAGCGGCCAAGTCTAGCGTCTCCCAGTCCAGTTGGTCTGGGACGCGTGAGAAGGGGA 5820

Qy       5821 GGGGCTTGGTGTAGAGGACCAGAAGACCC 5849
             ||||||||||||||||||||||||||||||
Db       5821 GGGGCTTGGTGTAGAGGACCAGAAGACCC 5849


RESULT 3
U89376/c
LOCUS       U89376                  1600 bp    DNA     linear   BCT 06-OCT-1997
DEFINITION  Thermus sp. YS45 replication protein (repT) gene, complete cds.
ACCESSION   U89376
VERSION     U89376.1  GI:2465516
KEYWORDS    .
SOURCE      Thermus sp. YS45
  ORGANISM  Thermus sp. YS45
            Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;
            Thermus.
REFERENCE   1  (bases 1 to 1600)
  AUTHORS   Wayne,J. and Xu,S.Y.
  TITLE     Identification of a thermophilic plasmid origin and its cloning
            within a new Thermus-E. coli shuttle vector
  JOURNAL   Gene 195 (2), 321-328 (1997)
  MEDLINE   97449309
   PUBMED   9305778
REFERENCE   2  (bases 1 to 1600)
  AUTHORS   Wayne,J. and Xu,S.-Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-FEB-1997) Research, New England Biolabs, 32 Tozer
            Road, Beverly, MA 01915, USA
FEATURES             Location/Qualifiers
     source          1. .1600
                     /organism="Thermus sp. YS45"
                     /mol_type="genomic DNA"
                     /strain="YS45"
                     /db_xref="taxon:64227"
     -35_signal      220. .225
                     /note="putative"
     -10_signal      243. .248
                     /note="putative"
     -35_signal      278. .283
                     /note="putative"
     -10_signal      302. .307
                     /note="putative"
     gene            332. .1369
                     /gene="repT"
     RBS             332. .336
                     /gene="repT"
     CDS             344. .1369
                     /gene="repT"
                     /function="required for plasmid replication"
                     /note="RepT"
                     /codon_start=1
                     /transl_table=11
                     /product="replication protein"
                     /protein_id="AAB72144.1"
                     /db_xref="GI:2465517"
                     /translation="MKNEKTFFEELYEALEETHDNTDATRGSDRGSEDFFLATDPPPD
                     GGAENRLAKGFTYQKEALRIALPEKDHEAFLSSVGAPPIPPAEPPVGNVCQAVQDGPQ
                     KLLELLQEIARSTIPYGNRELWRKVGTVVFMVPLEMLALNLGVTRQTVHAWKKVLEKK
                     GLVATDVLHQTVNGERRAIGTLWAVRLRPGKARLTLDDYIYPWRNLALDMANGVLSFN
                     WVKAYQDHGIRPTLDVLVLWAQGKRVMPNTKTVAVDLGLILVLPEVERSKLPALITLI
                     ATYIADLLDDRRSRRFYAGLLWAVARGELPAQYLFAVLMRVIRDYTDGHLTRPGAYLV
                     KTLKEAS"
     misc_feature    1158. .1166
                     /gene="repT"
                     /note="putative; encodes DnaA box"
     misc_feature    1293. .1301
                     /gene="repT"
                     /note="putative; encodes DnaA box"
ORIGIN

  Query Match           27.4%;  Score 1600;  DB 1;  Length 1600;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1600;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2901 CACGCACTCCAACCAAGGAATCCGCCAAAGGCGGCCTACCTTTTGAGCCCGTATCTTCCC 2960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1600 CACGCACTCCAACCAAGGAATCCGCCAAAGGCGGCCTACCTTTTGAGCCCGTATCTTCCC 1541

Qy       2961 CTGACGTATAGACCTTCGGATCGTCTCAGGGTGCACCCGAAGGATGTCTGCAAGCTCCTC 3020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1540 CTGACGTATAGACCTTCGGATCGTCTCAGGGTGCACCCGAAGGATGTCTGCAAGCTCCTC 1481

Qy       3021 GGGGGTCAGGTACACGGGCTTCATCCTCATGACACAACCTTACCCCACAGAGGACAACAC 3080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1480 GGGGGTCAGGTACACGGGCTTCATCCTCATGACACAACCTTACCCCACAGAGGACAACAC 1421

Qy       3081 ATGCAACTATGGGCAAAGTAGACAACGAGACCAAAAGCTTGGGCCACTCTCTCAGGAGGC 3140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1420 ATGCAACTATGGGCAAAGTAGACAACGAGACCAAAAGCTTGGGCCACTCTCTCAGGAGGC 1361

Qy       3141 CTCCTTGAGGGTCTTCACTAGGTACGCTCCCGGTCGTGTCAGATGGCCATCCGTGTAATC 3200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1360 CTCCTTGAGGGTCTTCACTAGGTACGCTCCCGGTCGTGTCAGATGGCCATCCGTGTAATC 1301

Qy       3201 TCGGATAACCCGCATTAGGACGGCAAATAGATATTGCGCGGGGAGTTCACCCCTGGCCAC 3260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1300 TCGGATAACCCGCATTAGGACGGCAAATAGATATTGCGCGGGGAGTTCACCCCTGGCCAC 1241

Qy       3261 AGCCCACAGCAAGCCTGCATAGAAACGTCTTGAACGACGGTCATCTAGGAGATCGGCAAT 3320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1240 AGCCCACAGCAAGCCTGCATAGAAACGTCTTGAACGACGGTCATCTAGGAGATCGGCAAT 1181

Qy       3321 GTACGTAGCAATGAGGGTGATAAGGGCCGGGAGTTTGGAACGCTCCACCTCGGGGAGGAC 3380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1180 GTACGTAGCAATGAGGGTGATAAGGGCCGGGAGTTTGGAACGCTCCACCTCGGGGAGGAC 1121

Qy       3381 CAGGATGAGGCCCAGGTCAACGGCCACGGTCTTGGTGTTGGGCATCACCCTTTTCCCCTG 3440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1120 CAGGATGAGGCCCAGGTCAACGGCCACGGTCTTGGTGTTGGGCATCACCCTTTTCCCCTG 1061

Qy       3441 AGCCCAGAGGACCAGCACGTCCAGGGTGGGGCGGATTCCGTGGTCCTGGTAGGCCTTGAC 3500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1060 AGCCCAGAGGACCAGCACGTCCAGGGTGGGGCGGATTCCGTGGTCCTGGTAGGCCTTGAC 1001

Qy       3501 CCAGTTGAAGGAGAGCACGCCGTTGGCCATGTCTAGGGCGAGGTTCCTCCAGGGGTAGAT 3560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1000 CCAGTTGAAGGAGAGCACGCCGTTGGCCATGTCTAGGGCGAGGTTCCTCCAGGGGTAGAT 941

Qy       3561 GTAGTCGTCCAGGGTGAGCCTGGCTTTCCCTGGCCTCAGCCGGACGGCCCAAAGGGTGCC 3620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        940 GTAGTCGTCCAGGGTGAGCCTGGCTTTCCCTGGCCTCAGCCGGACGGCCCAAAGGGTGCC 881

Qy       3621 GATGGCCCGGCGCTCCCCGTTGACGGTTTGGTGAAGGACGTCGGTGGCCACCAGGCCCTT 3680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        880 GATGGCCCGGCGCTCCCCGTTGACGGTTTGGTGAAGGACGTCGGTGGCCACCAGGCCCTT 821

Qy       3681 TTTCTCAAGGACCTTCTTCCAGGCGTGGACGGTCTGCCGGGTGACCCCCAGGTTGAGGGC 3740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        820 TTTCTCAAGGACCTTCTTCCAGGCGTGGACGGTCTGCCGGGTGACCCCCAGGTTGAGGGC 761

Qy       3741 CAACATCTCCAGGGGGACCATGAAGACGACCGTCCCCACCTTCCTCCAGAGCTCCCGGTT 3800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        760 CAACATCTCCAGGGGGACCATGAAGACGACCGTCCCCACCTTCCTCCAGAGCTCCCGGTT 701

Qy       3801 GCCGTAGGGGATGGTGGAGCGGGCAATCTCCTGGAGGAGTTCCAGAAGCTTCTGAGGCCC 3860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        700 GCCGTAGGGGATGGTGGAGCGGGCAATCTCCTGGAGGAGTTCCAGAAGCTTCTGAGGCCC 641

Qy       3861 GTCCTGGACGGCTTGACATACATTCCCAACGGGGGGTTCAGCTGGTGGTATAGGGGGGGC 3920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        640 GTCCTGGACGGCTTGACATACATTCCCAACGGGGGGTTCAGCTGGTGGTATAGGGGGGGC 581

Qy       3921 CCCAACAGAGGAAAGGAAAGCCTCATGGTCTTTCTCGGGTAAAGCAATCCTAAGTGCCTC 3980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        580 CCCAACAGAGGAAAGGAAAGCCTCATGGTCTTTCTCGGGTAAAGCAATCCTAAGTGCCTC 521

Qy       3981 TTTTTGGTATGTAAAGCCCTTCGCGAGGCGATTTTCGGCACCTCCATCTGGAGGGGGGTC 4040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        520 TTTTTGGTATGTAAAGCCCTTCGCGAGGCGATTTTCGGCACCTCCATCTGGAGGGGGGTC 461

Qy       4041 GGTGGCCAAGAAGAAGTCCTCTGACCCCCTATCTGACCCCCTAGTGGCATCGGTGTTGTC 4100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        460 GGTGGCCAAGAAGAAGTCCTCTGACCCCCTATCTGACCCCCTAGTGGCATCGGTGTTGTC 401

Qy       4101 GTGGGTTTCCTCTAAAGCCTCGTAAAGCTCTTCAAAGAAGGTTTTTTCGTTCTTCACCCT 4160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        400 GTGGGTTTCCTCTAAAGCCTCGTAAAGCTCTTCAAAGAAGGTTTTTTCGTTCTTCACCCT 341

Qy       4161 CGGACCTCCTTGTCATCTGGAGCCCGAGGCGTTACCCTAGGTCTTGGGGGTGATCCGGGG 4220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        340 CGGACCTCCTTGTCATCTGGAGCCCGAGGCGTTACCCTAGGTCTTGGGGGTGATCCGGGG 281

Qy       4221 CAACCGCCTCGGTTTCGCCTTTTTATGGGTCCAAAATAACCGTCAGCCCAGCGGCTGGCA 4280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        280 CAACCGCCTCGGTTTCGCCTTTTTATGGGTCCAAAATAACCGTCAGCCCAGCGGCTGGCA 221

Qy       4281 ATCCCCCCTCCTAAAAGGCCGTTATAGGCCCTGCTAGGAGGGGGGTAGTACTTTCCTACC 4340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        220 ATCCCCCCTCCTAAAAGGCCGTTATAGGCCCTGCTAGGAGGGGGGTAGTACTTTCCTACC 161

Qy       4341 CCCCTAGGCTTGGAGAGGCCTTAGGAGGTCTCCTAGGGCCTCGTGGGGGTGTAGGGGTAA 4400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        160 CCCCTAGGCTTGGAGAGGCCTTAGGAGGTCTCCTAGGGCCTCGTGGGGGTGTAGGGGTAA 101

Qy       4401 CCTCATGGCCAGGCCGGCCGGCTCGGGACTCTGGAGGAGGCCTCCATAGCCTACTCGTGG 4460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        100 CCTCATGGCCAGGCCGGCCGGCTCGGGACTCTGGAGGAGGCCTCCATAGCCTACTCGTGG 41

Qy       4461 TGGAGGTTTGTGAAGGGGTTCACTAATGCATACGGCTAGC 4500
              |||||||||||||||||||||||||||||||||||||||
Db         40 TGGAGGTTTGTGAAGGGGTTCACTAATGCATACGGCTAGC 1
```

RESULT 4
AR031809/c

```
LOCUS       AR031809                1242 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 1 from patent US 5866422.
ACCESSION   AR031809
VERSION     AR031809.1  GI:5946098
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1242)
  AUTHORS   Wayne,J. and Xu,S.-y.
  TITLE     Method for cloning and producing the Tsp45I restriction
            endonuclease in E. coli
  JOURNAL   Patent: US 5866422-A 1 02-FEB-1999;
FEATURES             Location/Qualifiers
     source          1. .1242
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN
```

Query Match 21.2%; Score 1242; DB 6; Length 1242;
Best Local Similarity 100.0%; Pred. No. 4.3e-256;
Matches 1242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1609 CTAGAAGGCGGACACAATCTCAAACTTGTGCTGTAGCCTGGGGAAATCCTCTAACACCCT 1668
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1242 CTAGAAGGCGGACACAATCTCAAACTTGTGCTGTAGCCTGGGGAAATCCTCTAACACCCT 1183

Qy       1669 TCTAGTGAAGGCTTTGACCGCCTCCCAGGAGGCATCTATGCCGATGGATCGCCGCTTTAA 1728
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1182 TCTAGTGAAGGCTTTGACCGCCTCCCAGGAGGCATCTATGCCGATGGATCGCCGCTTTAA 1123

Qy       1729 GAGGGGTGAGGCTATAAGCGTAGTACCGGAGCCTGCGAAGGGATCGAGCACTAAATCCCC 1788
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1122 GAGGGGTGAGGCTATAAGCGTAGTACCGGAGCCTGCGAAGGGATCGAGCACTAAATCCCC 1063

Qy       1789 CTCGTTACTCCCTGTTTGGACGATGAGCTTGAGCATGTCCAGATTTTTCTCGGTGGGGTA 1848
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1062 CTCGTTACTCCCTGTTTGGACGATGAGCTTGAGCATGTCCAGATTTTTCTCGGTGGGGTA 1003

Qy       1849 TCGCGGGTACGGAGGATCCTTGAACTGCCAAACGTCCTGGAGCTTCTTCCCCTTCTTCAG 1908
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1002 TCGCGGGTACGGAGGATCCTTGAACTGCCAAACGTCCTGGAGCTTCTTCCCCTTCTTCAG 943

Qy       1909 GCGATCCCGAGCGTAAACTTTCTTCCGCGGCACCCCGTTCTTTGACCAGACAATAAGCCC 1968
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        942 GCGATCCCGAGCGTAAACTTTCTTCCGCGGCACCCCGTTCTTTGACCAGACAATAAGCCC 883

Qy       1969 TTGAGCGTCTAGCTCGTCAAGCTTCTCCGGGGGATAGCGCCAATGCCGTCCAGGAGGGGG 2028
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        882 TTGAGCGTCTAGCTCGTCAAGCTTCTCCGGGGGATAGCGCCAATGCCGTCCAGGAGGGGG 823

Qy       2029 AAGTATTCCTCGCCAAGGCCTTCCGGTAGGGCCATCCTTGGTTTCTCCAGGAGCATGCAG 2088
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        822 AAGTATTCCTCGCCAAGGCCTTCCGGTAGGGCCATCCTTGGTTTCTCCAGGAGCATGCAG 763

Qy       2089 GGGATTGGTGGTGTACCGTTCCCCGTTCTCGTCTACAAAGGGGAAAAGCCTAGCGATCTC 2148
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        762 GGGATTGGTGGTGTACCGTTCCCCGTTCTCGTCTACAAAGGGGAAAAGCCTAGCGATCTC 703

Qy       2149 CTCTTCCGAATAGGGGCTAGCCGATTCGTTCCAAACGTAGTCCCGCGTTTTGGAGTAGAC 2208
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        702 CTCTTCCGAATAGGGGCTAGCCGATTCGTTCCAAACGTAGTCCCGCGTTTTGGAGTAGAC 643

Qy       2209 GAGGATCATGTCCTTTTGCGATCCGAAGGCCTTACGGGAAAAGTTTTTGGGATTTGAAGC 2268
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        642 GAGGATCATGTCCTTTTGCGATCCGAAGGCCTTACGGGAAAAGTTTTTGGGATTTGAAGC 583

Qy       2269 GATGCGGGCGATATGGTTAACGAAGTTTCGCCGGCCAAAGACCTCATCAAGGATGAGCTT 2328
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        582 GATGCGGGCGATATGGTTAACGAAGTTTCGCCGGCCAAAGACCTCATCAAGGATGAGCTT 523

Qy       2329 CACCTCGAACCCGTATTTCTCGTCTATGTGAACGAAGATCAGTCCTGAGTCCGCCATCAG 2388
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        522 CACCTCGAACCCGTATTTCTCGTCTATGTGAACGAAGATCAGTCCTGAGTCCGCCATCAG 463

Qy       2389 CTCCCTGAGAAGTATCAAGCGCTCCCTCAGGAACTCCACAAACTGAGGACCATCGAGGGT 2448
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        462 CTCCCTGAGAAGTATCAAGCGCTCCCTCAGGAACTCCACAAACTGAGGACCATCGAGGGT 403

Qy       2449 GTCATCGTAGCCCAACTGACCGTTTTTGGGCTGGCTGACGGTAGCAACGCGATCTGTTTC 2508
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        402 GTCATCGTAGCCCAACTGACCGTTTTTGGGCTGGCTGACGGTAGCAACGCGATCTGTTTC 343

Qy       2509 ATCGCCGCCAACGAGAAACTGCTGGCCGGTTCCATAAGGCGGGTCAATATAGACCAACTG 2568
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        342 ATCGCCGCCAACGAGAAACTGCTGGCCGGTTCCATAAGGCGGGTCAATATAGACCAACTG 283

Qy       2569 GACCTTCCCCGCATACCCACCAGGCTCCCGGAGCATCCACCGGAGAACCTGACCGTTTTC 2628
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        282 GACCTTCCCCGCATACCCACCAGGCTCCCGGAGCATCCACCGGAGAACCTGACCGTTTTC 223

Qy       2629 CCCCAAAAAGTAGGTGCCAATAGGATCAATCTCAAAAAGGGGGGCATTTCCCCCTAGGAA 2688
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        222 CCCCAAAAAGTAGGTGCCAATAGGATCAATCTCAAAAAGGGGGGCATTTCCCCCTAGGAA 163

Qy       2689 GAGGAGGGTTTCTTTTCGCAAAACAAGTTGTGGGGTGGGCTGATCAAGAATCTCCTTCTC 2748
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        162 GAGGAGGGTTTCTTTTCGCAAAACAAGTTGTGGGGTGGGCTGATCAAGAATCTCCTTCTC 103

Qy       2749 ATCGCGTTTTCCGGGGTAGACCAACCTAAAGGGCGAAGGTTCCGAGGTTTTCGAGGCTTT 2808
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        102 ATCGCGTTTTCCGGGGTAGACCAACCTAAAGGGCGAAGGTTCCGAGGTTTTCGAGGCTTT 43

Qy       2809 CAAGGGGGCTTTTCGGGTCAAACCAGGGTAGCTACGGCTCAT 2850
             ||||||||||||||||||||||||||||||||||||||||||
Db          42 CAAGGGGGCTTTTCGGGTCAAACCAGGGTAGCTACGGCTCAT 1
```

RESULT 5
AR139424/c

LOCUS	AR139424	1026 bp	DNA	linear	PAT 16-JUN-2001
DEFINITION	Sequence 4 from patent US 6207377.				
ACCESSION	AR139424				
VERSION	AR139424.1 GI:14481920				
KEYWORDS	.				
SOURCE	Unknown.				
ORGANISM	Unknown.				
	Unclassified.				
REFERENCE	1 (bases 1 to 1026)				
AUTHORS	Wayne,J. and Xu,S.-y.				
TITLE	Method for construction of thermus-E. coli shuttle vectors and identification of two Thermus plasmid replication origins				
JOURNAL	Patent: US 6207377-A 4 27-MAR-2001;				

FEATURES Location/Qualifiers
source 1. .1026
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN

Query Match 17.5%; Score 1026; DB 6; Length 1026;
Best Local Similarity 100.0%; Pred. No. 1.2e-209;
Matches 1026; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       3132 TCAGGAGGCCTCCTTGAGGGTCTTCACTAGGTACGCTCCCGGTCGTGTCAGATGGCCATC 3191
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1026 TCAGGAGGCCTCCTTGAGGGTCTTCACTAGGTACGCTCCCGGTCGTGTCAGATGGCCATC 967

Qy       3192 CGTGTAATCTCGGATAACCCGCATTAGGACGGCAAATAGATATTGCGCGGGGAGTTCACC 3251
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        966 CGTGTAATCTCGGATAACCCGCATTAGGACGGCAAATAGATATTGCGCGGGGAGTTCACC 907

Qy       3252 CCTGGCCACAGCCCACAGCAAGCCTGCATAGAAACGTCTTGAACGACGGTCATCTAGGAG 3311
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        906 CCTGGCCACAGCCCACAGCAAGCCTGCATAGAAACGTCTTGAACGACGGTCATCTAGGAG 847

Qy       3312 ATCGGCAATGTACGTAGCAATGAGGGTGATAAGGGCCGGGAGTTTGGAACGCTCCACCTC 3371
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        846 ATCGGCAATGTACGTAGCAATGAGGGTGATAAGGGCCGGGAGTTTGGAACGCTCCACCTC 787

Qy       3372 GGGGAGGACCAGGATGAGGCCCAGGTCAACGGCCACGGTCTTGGTGTTGGGCATCACCCT 3431
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        786 GGGGAGGACCAGGATGAGGCCCAGGTCAACGGCCACGGTCTTGGTGTTGGGCATCACCCT 727

Qy       3432 TTTCCCCTGAGCCCAGAGGACCAGCACGTCCAGGGTGGGGCGGATTCCGTGGTCCTGGTA 3491
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        726 TTTCCCCTGAGCCCAGAGGACCAGCACGTCCAGGGTGGGGCGGATTCCGTGGTCCTGGTA 667

Qy       3492 GGCCTTGACCCAGTTGAAGGAGAGCACGCCGTTGGCCATGTCTAGGGCGAGGTTCCTCCA 3551
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        666 GGCCTTGACCCAGTTGAAGGAGAGCACGCCGTTGGCCATGTCTAGGGCGAGGTTCCTCCA 607

Qy       3552 GGGGTAGATGTAGTCGTCCAGGGTGAGCCTGGCTTTCCCTGGCCTCAGCCGGACGGCCCA 3611
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        606 GGGGTAGATGTAGTCGTCCAGGGTGAGCCTGGCTTTCCCTGGCCTCAGCCGGACGGCCCA 547

Qy       3612 AAGGGTGCCGATGGCCCGGCGCTCCCCGTTGACGGTTTGGTGAAGGACGTCGGTGGCCAC 3671
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        546 AAGGGTGCCGATGGCCCGGCGCTCCCCGTTGACGGTTTGGTGAAGGACGTCGGTGGCCAC 487

Qy       3672 CAGGCCCTTTTTCTCAAGGACCTTCTTCCAGGCGTGGACGGTCTGCCGGGTGACCCCCAG 3731
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        486 CAGGCCCTTTTTCTCAAGGACCTTCTTCCAGGCGTGGACGGTCTGCCGGGTGACCCCCAG 427

Qy       3732 GTTGAGGGCCAACATCTCCAGGGGGACCATGAAGACGACCGTCCCCACCTTCCTCCAGAG 3791
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        426 GTTGAGGGCCAACATCTCCAGGGGGACCATGAAGACGACCGTCCCCACCTTCCTCCAGAG 367

Qy       3792 CTCCCGGTTGCCGTAGGGGATGGTGGAGCGGGCAATCTCCTGGAGGAGTTCCAGAAGCTT 3851
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        366 CTCCCGGTTGCCGTAGGGGATGGTGGAGCGGGCAATCTCCTGGAGGAGTTCCAGAAGCTT 307

Qy       3852 CTGAGGCCCGTCCTGGACGGCTTGACATACATTCCCAACGGGGGGTTCAGCTGGTGGTAT 3911
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        306 CTGAGGCCCGTCCTGGACGGCTTGACATACATTCCCAACGGGGGGTTCAGCTGGTGGTAT 247

Qy       3912 AGGGGGGGCCCCAACAGAGGAAAGGAAAGCCTCATGGTCTTTCTCGGGTAAAGCAATCCT 3971
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        246 AGGGGGGGCCCCAACAGAGGAAAGGAAAGCCTCATGGTCTTTCTCGGGTAAAGCAATCCT 187

Qy       3972 AAGTGCCTCTTTTTGGTATGTAAAGCCCTTCGCGAGGCGATTTTCGGCACCTCCATCTGG 4031
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        186 AAGTGCCTCTTTTTGGTATGTAAAGCCCTTCGCGAGGCGATTTTCGGCACCTCCATCTGG 127

Qy       4032 AGGGGGGTCGGTGGCCAAGAAGAAGTCCTCTGACCCCCTATCTGACCCCCTAGTGGCATC 4091
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        126 AGGGGGGTCGGTGGCCAAGAAGAAGTCCTCTGACCCCCTATCTGACCCCCTAGTGGCATC 67

Qy       4092 GGTGTTGTCGTGGGTTTCCTCTAAAGCCTCGTAAAGCTCTTCAAAGAAGGTTTTTTCGTT 4151
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 GGTGTTGTCGTGGGTTTCCTCTAAAGCCTCGTAAAGCTCTTCAAAGAAGGTTTTTTCGTT 7

Qy       4152 CTTCAC 4157
             ||||||
Db          6 CTTCAC 1
```

RESULT 6
AR031810

LOCUS	AR031810	999 bp	DNA	linear	PAT 29-SEP-1999
DEFINITION	Sequence 3 from patent US 5866422.				
ACCESSION	AR031810				
VERSION	AR031810.1 GI:5946099				
KEYWORDS	.				
SOURCE	Unknown.				
ORGANISM	Unknown.				
	Unclassified.				
REFERENCE	1 (bases 1 to 999)				
AUTHORS	Wayne,J. and Xu,S.-y.				
TITLE	Method for cloning and producing the Tsp45I restriction				

```
          endonuclease in E. coli
  JOURNAL   Patent: US 5866422-A 3 02-FEB-1999;
FEATURES             Location/Qualifiers
     source          1. .999
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             17.1%;  Score 999;  DB 6;  Length 999;
  Best Local Similarity   100.0%;  Pred. No. 7.6e-204;
  Matches  999;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        614 ATGCAACAGATGGCCGAGTGGAACGTGTGGACACAGAGAAGCGTTGAGCTTCTGGAGAAG 673
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCAACAGATGGCCGAGTGGAACGTGTGGACACAGAGAAGCGTTGAGCTTCTGGAGAAG 60

Qy        674 GGGTATTTGGATAAACTACTGCAGGTCTATAAAGGGGAAAGTGGCTCTTCGAGGTCAGTA 733
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGGTATTTGGATAAACTACTGCAGGTCTATAAAGGGGAAAGTGGCTCTTCGAGGTCAGTA 120

Qy        734 CCAGAGGAGGTAGAGGAAAAACTTCGCGAGGCCTACAAGGCATACGAGGGGAGGCAGGAT 793
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCAGAGGAGGTAGAGGAAAAACTTCGCGAGGCCTACAAGGCATACGAGGGGAGGCAGGAT 180

Qy        794 AGTCCGGAGGCAGAAACGAAACTCGTGGAAGCCGTGCTAAATGCCAGAAAAAAGGTCGAG 853
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGTCCGGAGGCAGAAACGAAACTCGTGGAAGCCGTGCTAAATGCCAGAAAAAAGGTCGAG 240

Qy        854 CGGTCCCCCTTCAATCACCCCTACCTGCCTTTGGTCTACTACCTGGTTTCGGAAAAAGCA 913
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CGGTCCCCCTTCAATCACCCCTACCTGCCTTTGGTCTACTACCTGGTTTCGGAAAAAGCA 300

Qy        914 GAAAAAGCGAACAAGGCCCTTGAGGAGGCATTGCAGGAGGTTGCCTCAAAGCACCCAGAA 973
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAAAAAGCGAACAAGGCCCTTGAGGAGGCATTGCAGGAGGTTGCCTCAAAGCACCCAGAA 360

Qy        974 ACCATCCGCGTCCTGGCCAAGGAAGCGCAAAGAAGAGGCGTAGAAGCCTTGATCCAAAGG 1033
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ACCATCCGCGTCCTGGCCAAGGAAGCGCAAAGAAGAGGCGTAGAAGCCTTGATCCAAAGG 420

Qy       1034 CTCAAGGAGCCTCCCGAAATAAATCGGCAGATAGGGCCGATGTTCAAAAGGTGGTACAAA 1093
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTCAAGGAGCCTCCCGAAATAAATCGGCAGATAGGGCCGATGTTCAAAAGGTGGTACAAA 480

Qy       1094 GAAGAGCTAAAGGGGAAAATAGAAGAGAGGCTTCCAGGCCCTACCAAACCAAAGATTGTG 1153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GAAGAGCTAAAGGGGAAAATAGAAGAGAGGCTTCCAGGCCCTACCAAACCAAAGATTGTG 540

Qy       1154 GTAGTATCCCCTGAAAAAAGTAAACCGGAGCAAGCACCCCTTATTGCGGAGAGAGAAGCG 1213
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTAGTATCCCCTGAAAAAAGTAAACCGGAGCAAGCACCCCTTATTGCGGAGAGAGAAGCG 600

Qy       1214 GGCATCATCATATACACGGGATCGGATGAAGCTTTGAAAGATGCCGCCAAGGAAAACCTG 1273
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCATCATCATATACACGGGATCGGATGAAGCTTTGAAAGATGCCGCCAAGGAAAACCTG 660

Qy       1274 GGCCTTGGCGAGGAAGCAGAACTAGGCACCAAGGGCGTAGATTTCTACGTGGTCATCCGG 1333
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GGCCTTGGCGAGGAAGCAGAACTAGGCACCAAGGGCGTAGATTTCTACGTGGTCATCCGG 720

Qy       1334 CGTAGCCCTGAAGAGACATGGCACCTAACAGGAGAAGTGAAGTTTCAATCCGACTTTGGC 1393
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CGTAGCCCTGAAGAGACATGGCACCTAACAGGAGAAGTGAAGTTTCAATCCGACTTTGGC 780

Qy       1394 GGAAACCAAGACAACCAGAAACTAGTAGCAAAGGCTTCCATAAGGTTGGACCTTGAGAAG 1453
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GGAAACCAAGACAACCAGAAACTAGTAGCAAAGGCTTCCATAAGGTTGGACCTTGAGAAG 840

Qy       1454 AGGCACATAGGAATAGTGGTGGTGGACGGAATGCCTGTGGTGAGCAAGTTTCGTGGGTGG 1513
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AGGCACATAGGAATAGTGGTGGTGGACGGAATGCCTGTGGTGAGCAAGTTTCGTGGGTGG 900

Qy       1514 GCCGGACTGGGGAAAGAAACGATCGTTACATCCGTACTCCTCCTTCCAGACCTGATAGCG 1573
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GCCGGACTGGGGAAAGAAACGATCGTTACATCCGTACTCCTCCTTCCAGACCTGATAGCG 960

Qy       1574 GAGCTCTACCAAAAGGGTGAAGAAGCCCTGGGCCTCTAG 1612
              ||||||||||||||||||||||||||||||||||||||||
Db        961 GAGCTCTACCAAAAGGGTGAAGAAGCCCTGGGCCTCTAG 999


RESULT 7
AF328924
LOCUS       AF328924                4938 bp    DNA     linear   BCT 12-NOV-2002
DEFINITION  Helicobacter pylori strain CR38 cytosine-specific DNA
            methyltransferase, cytosine-specific DNA methyltransferase,
            putative HP0052-like protein, type II DNA modification enzyme, and
            type II restriction enzyme genes, complete cds.
ACCESSION   AF328924
VERSION     AF328924.1  GI:17225546
KEYWORDS    .
SOURCE      Helicobacter pylori
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
            Helicobacteraceae; Helicobacter.
REFERENCE   1  (bases 1 to 4938)
  AUTHORS   Chanto,G., Occhialini,A., Gras,N., Alm,R.A., Megraud,F. and
            Marais,A.
  TITLE     Identification of strain-specific genes located outside the
            plasticity zone in nine clinical isolates of Helicobacter pylori
  JOURNAL   Microbiology 148 (Pt 11), 3671-3680 (2002)
   PUBMED   12427957
REFERENCE   2  (bases 1 to 4938)
  AUTHORS   Chanto,G., Occhialini,A., Gras,N., Megraud,F. and Marais,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-DEC-2000) Laboratoire de Bacteriologie, Universite
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .4938
                     /organism="Helicobacter pylori"
                     /mol_type="genomic DNA"
                     /strain="CR38"
                     /specific_host="Homo sapiens"
                     /db_xref="taxon:210"
                     /map="region F; between JHP43/HP50 and JHP47/HP55"
                     /note="isolated from patient with chronic gastritis"
     CDS             complement(1. .705)
                     /note="HP0051-like protein"
                     /codon_start=1
                     /transl_table=11
                     /product="cytosine-specific DNA methyltransferase"
                     /protein_id="AAL37468.1"
                     /db_xref="GI:17225547"
                     /translation="MKNLISCTKGYFLEEIKERLNALGYQLSYQILNAKDYGVPQNRE
                     RAFIVGASRFSFDFNLLEPSQSVSVQDAISDLAYLCSNEGAFESDYLNPIQSSYQALM
                     RKDSPKLYNHQATNHSQAALEKLKLINKEKGKECLPKNLHGKQQFKSTWGRLNWNKIS
                     PTIDTRFDTPSNGTNSHPELHRSITPREAARIQSFSDNYIFYGNKTSVCKQIGNAVPP
                     LLALALGKAILKSLRK"
     CDS             complement(750. .1073)
                     /note="HP0051-like protein"
                     /codon_start=1
                     /transl_table=11
                     /product="cytosine-specific DNA methyltransferase"
                     /protein_id="AAL37469.1"
                     /db_xref="GI:17225548"
                     /translation="MNYKILDLFCGAGGFSAGLEYLKEFDALIGLDCDKQALITFENN
                     HKNAIGVCGDITQTEIKEKVIELAKKLEINMIIGGPPCQGFYNKGKNLGLKDPRNFLF
                     LEYNI"
     CDS             complement(1070. .2305)
                     /codon_start=1
                     /transl_table=11
                     /product="putative HP0052-like protein"
                     /protein_id="AAL37470.1"
                     /db_xref="GI:17225549"
                     /translation="MLFVSRFLNDNIDPFNLGVLLSRFQIKNDRIYGVCSYKSSKFTP
                     GYEESKAQVLNALNTLSTHPIWQSNQESVTKIKGTFVFILENDLHLDENAFYKKLLNL
```

```
                     IIDNDFFNRSNLMNLMTLNQKRFLSGFFESRGSIDTQRNFLTLDYFFHSPLEFNKFHY
                     LIDFFHIPSEALNFNFRELQHEYAQGINQRNVQFRIYLNWYLCHIGLFNPYKARIAEH
                     VFKTTLIYDGVYYKLSYPPTTKYHGNGFIERAHFYLKNVYQQDLDKKRIKESLREQLG
                     WIQKSEEFKRDSKIISFYRLSTPNVCNACCDDYDIKERSFLSLPLYQITQRSDSYYTE
                     LHHVISLGKNNELDVLENQLCPTCHRALKKGASAEGFQKRLIGKILKRNKGNLEFVQL
                     RFETDDLPTLIDKIYESLK"
     CDS             complement(2663..3694)
                     /function="methyltransferase"
                     /note="JHP45-like protein"
                     /codon_start=1
                     /transl_table=11
                     /product="type II DNA modification enzyme"
                     /protein_id="AAL37471.1"
                     /db_xref="GI:17225550"
                     /translation="MQNLLIQAENAIALLFLLNDKNLKGKIDLIYIDPPFATNNHFTI
                     TNGRATTISNSKNGDIAYSDKVVGMDFIEFLKQRLVLLKELLSEQGSIYVHTDCKIGH
                     YVKVMLDEIFGIQNFINEITRIKCNPKNFKRMGYGNIKDMILFYSKGKNPIFNEPKIP
                     YTPQDLEKRFPKIDKDKRRYTTVPIHAPGEVESGECSKAFKGMLPPKGRHWRTDIATL
                     ERWDKEGLIEYSNNNNPRKKIYALEQVGKRVQDIWEFKDPQYPSYPTEKNAQLLDLII
                     KTSSNKDSIVLDCFCGSGTTLKSAFLLQRKFIGIDNSNLAIQACKNKLETITKDLFVS
                     QNFYDFLVF"
     CDS             complement(3704..4489)
                     /note="JHP46-like protein"
                     /codon_start=1
                     /transl_table=11
                     /product="type II restriction enzyme"
                     /protein_id="AAL37472.1"
                     /db_xref="GI:17225551"
                     /translation="MVNYWTKLSIEYANQRSYLDDLFQVYPTIPEGLREIDSKIWSNV
                     EYHFKRKDNLALITELLNLDLFPIKDSYMAYLKRDKSALERNPRTINRICGRLYGMGL
                     NKIFEKCSEPKETNRQIGPMFKDWINNKSLGVEPVDLNNFIANENDAILRASDNIMAE
                     FAKSHLNYHHHKGLDFVARFNKKYIIGEAKFLTDFGGHQNAQFNDAISTIEAPNIKAI
                     KVAILDGVLYIESNNKMRKLLDTTYRNYNIMSALVLREFLYQI"
ORIGIN

  Query Match             3.2%;  Score 187.2;  DB 1;  Length 4938;
  Best Local Similarity   52.1%;  Pred. No. 2.5e-29;
  Matches  476;  Conservative    0;  Mismatches  423;  Indels   15;  Gaps    2;

Qy       1664 ACCCTTCTAGTGAAGGCTTTGACCGCCTCCCAGGAGGCATCTATGCCGATGGATCGCCGC 1723
Db       2724 AGCTTGTTTTTGCAAGCTTGGATAGCTAAATTGGAATTATCAATGCCTATAAATTTTCGT 2783

Qy       1724 TTTAAGAGGGGTGAGGCTATAAGCGTAGTACCGGAGCCTGCGAAGGGATCGAGCACTAAA 1783
Db       2784 TGCAATAAAAACGCAGATTTTAAGGTTGTTCCAGAACCACAAAAACAATCTAAAACAATA 2843

Qy       1784 TCCCCCTCGTTACTCCCTGTTTGGACGATGAGCTTGAGCATGTCCAGATTTTTCTCGGTG 1843
Db       2844 CTATCTTTATTAGAAGAGGTTTTAATGATTAAGTCTAATAATTGAGCGTTTTTTTCTGTA 2903

Qy       1844 GGGTATCGCGGGTACGGAGGATCCTTGAACTGCCAAACGTCCTGGAGCTTCTTCCCCTTC 1903
Db       2904 GGGTAGCTTGGATATTGTGGGTCTTTAAATTCCCAGATGTCTTGGACTCTTTTGCC---- 2959

Qy       1904 TTCAGGCGATCCCGAGCGTAAACTTTCTTCCGCGGCACCCCGTTCTTTGACCAGACAATA 1963
Db       2960 --AACTTGTTCTAAGGCATAAATTTTTTTTCTAGGGTTATTATTGTTAGAGTATTCAATC 3017

Qy       1964 AGCCCTTGAGCGTCTAGCTCGTCAAGCTTCTCCGGGGGATAGCGCCAATGCCGTCCAGGA 2023
Db       3018 AAACCTTCTTTGTCCCAACGCTCAAGTGTGGCAATATCAGTGCGCCAATGCCGCCCTTTT 3077

Qy       2024 GGGGGAAGTATTCCTCGCCAAGGCCTTCCGGTAGGGCCATCCTTGGTTTCTCCAGGAGCA 2083
Db       3078 GGAGGTAGCATACCTTTAAATGCTTTAGAACATTCGCCACTTTCTACTTCTCCTGGAGCA 3137

Qy       2084 TGCAGGGGATTGGTGGTGTACCGTTCCCCGTTCTCGTCTACAAAGGGGAAAAGCCTAGCG 2143
Db       3138 TGTATTGGAACGGTAGTGTAACGCCTTTTATCCTTGTCAATTTAGGGAATCGTTTTTCT 3197

Qy       2144 ATCTCCTCTTCCGAATAGGGGCTAGCCGATTCGTTCCAAACGTAGTCCCGCGTTTTGGAG 2203
Db       3198 AAATCTTGTGGCGTATAAGGGATCTTAGGTTCGTTAAAAATGGGATTTTTTCCTTTAGAG 3257

Qy       2204 TAGACGAGGATCATGTCCTTTTGCGATCCGAAGGCCTTACGGGAAAAGTTTTTGGGATTT 2263
Db       3258 TAAAATAAAATCATATCTTTTATATTGCCATAGCCCATCCTTTTAAAATTTTTAGGATTG 3317

Qy       2264 GAAGCGATGCGGGCGATATGGTTAACGAAGTTTCGCCGGCCAAAGACCTCATCAAGGATG 2323
Db       3318 CACTTTATGCGTGTGATTTCATTTATAAAATTTTGTATGCCAAATATTTCATCTAACATC 3377

Qy       2324 AGCTTCACCTCGAACCCGTATTTCTCGTCTATGTGAACGAAGATCAGTCCTGAGTCCGCC 2383
Db       3378 ACCTTAACATAATGCCCTATCTTACAATCTGTATGCACATAGATAGAGCCTTGTTCTGAA 3437

Qy       2384 ATCAGCTCCCTGAGAAGTATCAAGCGCTCCCTCAGGAACTCCACAAACTGAGGACCATCG 2443
Db       3438 AGCAATTCTTTAAGCAATACCAGGCGTTGTTTTAAAAATTCTATAAAATCCATACCCACT 3497

Qy       2444 AGGGTGTCATCGTAGCCCAACTGACCGTTTTTGGGCTGGCTGACGGTAGCAACGCGATCT 2503
Db       3498 ACTTTGTCGCTATAAGCAATATCGCCATTCTTAGAATTGCTAATTGTGGTTGCTCTACCA 3557

Qy       2504 GTTTCATCGCCGCCAACGAGAAACTGCTGGCCGGTTCCATAAGGCGGGTCAATATAGACC 2563
Db       3558 TTT---------GTGATAGTAAAATGGTTGTTTGTAGCAAATGGAGGGTCAATATATATC 3608

Qy       2564 AACTGGACCTTCCC 2577
Db       3609 AAATCTATTTTTCC 3622


RESULT 8
AE001444
LOCUS       AE001444               11131 bp    DNA     linear   BCT 20-JAN-1999
DEFINITION  Helicobacter pylori, strain J99 section 5 of 132 of the complete
            genome.
ACCESSION   AE001444 AE001439
VERSION     AE001444.1  GI:4154549
KEYWORDS    .
SOURCE      Helicobacter pylori J99
  ORGANISM  Helicobacter pylori J99
            Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
            Helicobacteraceae; Helicobacter.
REFERENCE   1  (bases 1 to 11131)
  AUTHORS   Alm,R.A., Ling,L.S., Moir,D.T., King,B.L., Brown,E.D., Doig,P.C.,
            Smith,D.R., Noonan,B., Guild,B.C., deJonge,B.L., Carmel,G.,
            Tummino,P.J., Caruso,A., Uria-Nickelsen,M., Mills,D.M., Ives,C.,
            Gibson,R., Merberg,D., Mills,S.D., Jiang,Q., Taylor,D.E.,
            Vovis,G.F. and Trust,T.J.
  TITLE     Genomic-sequence comparison of two unrelated isolates of the human
            gastric pathogen Helicobacter pylori
  JOURNAL   Nature 397 (6715), 176-180 (1999)
  MEDLINE   99120557
   PUBMED   9923682
REFERENCE   2  (bases 1 to 11131)
  AUTHORS   King,B.L., Alm,R.A. and Trust,T.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JAN-1999) Astra Research Center Boston, 128 Sidney
            Street, Cambridge, MA 02139, USA
COMMENT     Address all correspondence to:  hp@arcb.us.astra.com  or  Richard
            A. Alm, Astra Research Center Boston, 128 Sidney Street, Cambridge,
            MA,  02139.  Lo-See L. Ling, Donald T. Moir, Douglas R. Smith,
            Braydon C. Guild, Gilles Carmel, Anthony Caruso, Debra M. Mills,
            Rene Gibson, and Gerald F. Vovis are with Genome Therapeutics
            Corporation, 100 Beaver Street, Waltham, MA, 02453.  Qin Jiang and
            Diane E. Taylor are with the University of Alberta Department of
            Medical Microbiology and Immunology, Edmonton, Alberta, Canada, T6G
            2H7 and the Canadian Bacterial Diseases Network.  All other authors
            are with Astra Research Center Boston, 128 Sidney Street,
            Cambridge, MA, 02139.  Putative identifications, sequence
            alignments, and name and sequence search capabiblity are available
            at ARCB's World Wide Web site. (URL:
            http://www.astra-boston.com/hpylori).
```

```
FEATURES             Location/Qualifiers
     source          1. .11131
                     /organism="Helicobacter pylori J99"
                     /mol_type="genomic DNA"
                     /strain="J99"
                     /db_xref="taxon:85963"
     gene            complement(248. .1279)
                     /gene="jhp0045"
     CDS             complement(248. .1279)
                     /gene="jhp0045"
                     /codon_start=1
                     /transl_table=11
                     /product="putative TYPE II DNA MODIFICATION ENZYME
                     (METHYLTRANSFERASE)"
                     /protein_id="AAD05627.1"
                     /db_xref="GI:4154550"
                     /translation="MQNLLIQAENAIALLFLLNDKNLKGKIDLIYIDPPFATNNHFTI
                     TNGRATTISNSKNGDIAYSDKVVGMDFMEFLKQRLVLLKELLSEQGSIYVHTDYKIGH
                     YVKVMLDEIFGIQNFRNEITRIKCNPKNFKRIGYGNIKDMILFYSKGKNPIFNEPKIP
                     YTPQDLEKRFPKIDKDKRRYTTVPIHAPGEVESGECSKAFKGMLPPKGRHWRTDIATL
                     ERWDKEGLIEYSNNNNPRKKIYALEQVGKRVQDIWEFKDPQYPSYPTEKNAQLLDLII
                     KTSSNKDSIVLDCFCGSGTTLKSAFLLQRKFIGIDNSDLAIQACKNKLETITKDLFVS
                     QNFYDFLVF"
     gene            complement(1289. .2071)
                     /gene="jhp0046"
     CDS             complement(1289. .2071)
                     /gene="jhp0046"
                     /codon_start=1
                     /transl_table=11
                     /product="putative TYPE II restriction enzyme"
                     /protein_id="AAD05628.1"
                     /db_xref="GI:4154551"
                     /translation="MNYWTKLSIEYANQRSYLDDLFQVYPTIPEGLREIDSKIWSNVE
                     YHFKRKDNLALITELLNLDLFPIKDSYMAYLKRDKSALERNPRTINRICGRLYEMGLN
                     KIFEKCSEPKETNRQIGPMFKDWINNKSLGVEPVDLNDFIANENDAILKASDNVMAEF
                     AKSHLNYHHHKGLDFVARFNKKYIIGEAKFLTDFGGHQNAQFNDAINTIETPNIKAIK
                     VAILDGVLYIESNNKMRKLLDTTYRNYHIMSALVLREFLYQI"
     gene            complement(2499. .3989)
                     /gene="putP"
                     /note="synonym: jhp0047"
     CDS             complement(2499. .3989)
                     /gene="putP"
                     /note="similar to H. pylori 26695 gene HP0055"
                     /codon_start=1
                     /transl_table=11
                     /product="sodium/proline symporter"
                     /protein_id="AAD05631.1"
                     /db_xref="GI:4154554"
                     /translation="MGHVVLSTPIVTMFVVYSLLMLYIGFYFYKQNETTEDYFLGDRS
                     MGPVISALSAGASDMSGWLLMGLPGALYVGGLINSHIAIGLSLGALINWVFVAKRLRI
                     YTSVIANSITISDYFETRFSDDKHILRLISAFVILIFFIFYISSGLVSGAKLFEATFG
                     IQYTYALSIGTLIIVSYTFLGGYKAVCWTDLIQGLLMMSALIVVPIVMIIHLGGIGEG
                     IKIIREIKPENLSFLQGSSVVAIISSLAWGLGYFGQPHILVRFMSIRSIRDVPKATTI
                     GISWMVISLIGACVMGLLGVAYVHKFDLSLEDPEKIFIVMSQLLFNPWITGILLSAIL
                     AAVMSTASSQLLVSSSTIAEDFYATIFNKNAPQKLVMTISRLSVLGVACIAFFISTDK
                     NASILSIVSYAWAGFGASFGSVILFSLFWSRMTRIGAIAGMLSGASTVILYDKFGKSF
                     LDIYEIVPGFIVASVAIVAFSLFSSVRSGTKEAFETMLKEIESLKH"
     gene            complement(4016. .7573)
                     /gene="putA"
                     /note="synonym: jhp0048"
     CDS             complement(4016. .7573)
                     /gene="putA"
                     /note="similar to H. pylori 26695 gene HP0056"
                     /codon_start=1
                     /transl_table=11
                     /product="Proline/pyrroline-5-carboxylate dehydrogenase"
                     /protein_id="AAD05632.1"
                     /db_xref="GI:4154555"
                     /translation="MQKIIEDSLELAKKLQDSISNHLSEQEKAFHFKMQKLLNNPENK
                     VMLIELMDRSFRCLDNKARFEMIEHVLDKYKSREIFSSFEKWLLMGFLSFGKMLPDMS
                     VPFFVNKIRSDTKAMVLDQEESQLRERILKRKNEKIILNVNFIGEEVLGEEEASARFE
                     KYSQALKSNYIQYISIKITTIFSQINILDFEYSKKEIVKRLDALYALALEEEKKQGMP
                     KFINLDMEEFRDLELTVESFMESIAKFDLNAGIVLQAYIPDSYEYLKKLHAFSKERVL
                     KGLKPIKIRFVKGANMESEETIASMKDWALPTFSSKQDTDSNYNKMLDFVLEGDNYKY
                     IHIGAASHNIFEIAYVYTRIHALNDPVVLEHFSFEMLEGMSLQASQELKEMHKLILYA
                     PVCDEAHFNNAIAYLVRRLDENTSSDNFMKAFFNLKVGTSEWKDQEQRFLNSLKGIAT
                     LDNTTHRTQDRNAKQTGHTTYPNHSFKNESDTDFILKANREWAKKVRDKMHNAPILEL
                     YPEIDGRFEDPNLTPLEVFDRIHHKKIASVHLADKEAILKALEVAKSDKSRFSQKSFT
                     EIHALLSQTAQLFRERRGDLIGISALEVGKTFAETDAEVSEAIDFLEFYPYSLRVLQE
                     QNKKTKFTPKGVGVVIAPWNFPVGISVGTIAAPLAAGNRVIYKPSSLSSVTGYKLCEC
                     FWDAGVPRDALIYLPSKGSDISEHLLKDESIKFAILTGGEDTAYKMLKANPTLALSAE
                     TGGKNATIVSKMADRDQAIKNVIHSAFSNSGQKCSATSLLVLEKEVYEDENFKKTLID
                     ATLSLSVGDPFDFKNKIGALADKPNEKVIKAIDELKSYENYEIPASFVDDNPYLMKPS
                     IKYGTKKGDFTHQTELFTPILSVMKAKDLDEAIEIVNSTGYGLTSALESLDEREWEYY
                     LERIEAGNIYINKPTTGAIVLRQPFGGVKKSAVGFGRKVGIFNYITQFVNIHQEEEDE
                     NALKNPLSEALEGLTQKGYDEHTHELKRAIFMAKSYAYHYKHEFSQAKDYVKIRGEDN
                     LFSYTKVTSVGYRITEKDTLSDMLGVALACLISQIPLTISIENERANKDLTFFLECLK
                     ALQANAPIVYESLQKFSEKLNAFNRVRYLKSDLDLLHEQASRLGMVLATAKPCLNGRF
                     ELLYYHLERSVSISYHRYGNLGSRVLRQPTCHKSCCAEK"
     gene            7894. .8103
                     /gene="jhp0049"
     CDS             7894. .8103
                     /gene="jhp0049"
                     /note="similar to H. pylori 26695 gene HP0057"
                     /codon_start=1
                     /transl_table=11
                     /product="putative"
                     /protein_id="AAD05633.1"
                     /db_xref="GI:4154556"
                     /translation="MKTIKNGIMIGTLGALLLSGCSSFDAQRFACLPKDHSSKDASTK
                     KEAQYIPKGFFDPYSSNLNHWDSTF"
     gene            8431. .9093
                     /gene="jhp0050"
     CDS             8431. .9093
                     /gene="jhp0050"
                     /note="similar to H. pylori 26695 gene HP0058"
                     /codon_start=1
                     /transl_table=11
                     /product="putative"
                     /protein_id="AAD05634.1"
                     /db_xref="GI:4154557"
                     /translation="MFGKSSNDTQENPKDALKSTNFHEKHTKPTETTELVEENKALTT
                     EKERLERENKNLTADKENLTKEKTELQKQVNELKNSKQVLENEKADWLREKENLTKDR
                     ENLTKEKTELTEKNKVLTTEKERLATEKENLTKEKTESQKQVNELKNSKQVLENEKAD
                     LTNENTKLKTDKTDLTEKNQRLTTEKTELNNKITGLATEKERLAADKENLTKESRQRK
                     PN"
     gene            9223. .9585
                     /gene="jhp0051"
     CDS             9223. .9585
                     /gene="jhp0051"
                     /note="similar to H. pylori 26695 gene HP0058"
                     /codon_start=1
                     /transl_table=11
                     /product="putative"
                     /protein_id="AAD05635.1"
                     /db_xref="GI:4154558"
                     /translation="MEVKDRLNFNFVATTHSAMDLIASVLSDSKYYLESFYNQASQEL
                     GDKRSDKGEKLAELFDLLFEYIKDSKFERLKEPSAYDYSCKKLYPEQNTSQKMRRVVL
                     RGYKHNDKMYHTIVDMGS"
     gene            9586. .10569
                     /gene="jhp0052"
     CDS             9586. .10569
                     /gene="jhp0052"
                     /note="similar to H. pylori 26695 gene HP0059"
                     /codon_start=1
                     /transl_table=11
                     /product="putative"
                     /protein_id="AAD05629.1"
                     /db_xref="GI:4154552"
                     /translation="MGTLIEKWFGFSQIREELEARISELEDENTELLREREYLAAETS
                     ELKDANDQLRQKNDKLFITKDKLTKENTELFAENESLSVKISGLEHSNDQLWQNNNKL
                     TKEKAELKTEKDILAKENTRLLAARDRLTEEKRELTTEKERLKRENTELTHKITELTK
                     ENKALTTENDKLNHQVTALTNERDSLEQERARLQDAHGFLEKRCTNLEKENQRLTDKL
                     KQLESAQKSLENTNNQLRQALENSNVQLAQAKEKIAIEKSELEREIARLKSLEGMEAK
                     SDLDLHNRRLASANEDLKRQNRKLEEENIALKERVDGLNEQLSKLQPQKPQ"
     gene            10580. .11011
```

```
                /gene="jhp0053"
     CDS        10580. .11011
                /gene="jhp0053"
                /note="similar to H. pylori 26695 gene HP0060"
                /codon_start=1
                /transl_table=11
                /product="putative"
                /protein_id="AAD05630.1"
                /db_xref="GI:4154553"
                /translation="MVKEKSLELPLGHPLVEKLCELSLNNKAAFNEKSEPNFKDEVSE
                EDKIKFKQALRVFHAIVNNETSLRYLSDENQKFLEDLAQAKKITNEQIEKTLEIVSDS
                DVDVDFEKLKKLMLNVDSVAVGLKSYSQNQLLDLDGGHWDL"
ORIGIN

  Query Match             3.2%;  Score 185.6;  DB 1;  Length 11131;
  Best Local Similarity   52.0%;  Pred. No. 5.4e-29;
  Matches  475;  Conservative    0;  Mismatches  424;  Indels   15;  Gaps    2;

Qy       1664 ACCCTTCTAGTGAAGGCTTTGACCGCCTCCCAGGAGGCATCTATGCCGATGGATCGCCGC 1723
Db        309 AGCTTGTTTTTGCAAGCTTGGATAGCCAAATCGGAATTATCAATGCCTATAAATTTTCGT 368

Qy       1724 TTTAAGAGGGGTGAGGCTATAAGCGTAGTACCGGAGCCTGCGAAGGGATCGAGCACTAAA 1783
Db        369 TGCAATAAAAACGCAGATTTTAAGGTTGTTCCAGAACCACAAAAACAATCTAAAACAATA 428

Qy       1784 TCCCCCTCGTTACTCCCTGTTTGGACGATGAGCTTGAGCATGTCCAGATTTTTCTCGGTG 1843
Db        429 CTATCTTTATTAGAAGAGGTTTTAATGATTAAGTCTAATAATTGAGCGTTTTTTTCTGTA 488

Qy       1844 GGGTATCGCGGGTACGGAGGATCCTTGAACTGCCAAACGTCCTGGAGCTTCTTCCCCTTC 1903
Db        489 GGGTAGCTTGGATATTGTGGGTCTTTAAATTCCCAGATGTCTTGGACTCTTTTGCC---- 544

Qy       1904 TTCAGGCGATCCCGAGCGTAAACTTTCTTCCGCGGCACCCCGTTCTTTGACCAGACAATA 1963
Db        545 --AACTTGTTCTAAGGCATAAATTTTTTTTCTAGGGTTATTATTGTTAGAATACTCAATC 602

Qy       1964 AGCCCTTGAGCGTCTAGCTCGTCAAGCTTCTCCGGGGGATAGCGCCAATGCCGTCCAGGA 2023
Db        603 AAACCTTCTTTATCCCAACGCTCAAGTGTGGCAATATCAGTGCGCCAATGCCGCCCTTTT 662

Qy       2024 GGGGGAAGTATTCCTCGCCAAGGCCTTCCGGTAGGGCCATCCTTGGTTTCTCCAGGAGCA 2083
Db        663 GGAGGTAGCATACCTTTAAATGCTTTAGAACATTCGCCACTTTCCACTTCTCCTGGAGCA 722

Qy       2084 TGCAGGGGATTGGTGGTGTACCGTTCCCCGTTCTCGTCTACAAAGGGGAAAAGCCTAGCG 2143
Db        723 TGTATTGGAACGGTAGTGTAACGCCTTTTATCTTTGTCAATTTTAGGGAATCGTTTTTCT 782

Qy       2144 ATCTCCTCTTCCGAATAGGGGCTAGCCGATTCGTTCCAAACGTAGTCCCGCGTTTTGGAG 2203
Db        783 AAATCTTGTGGCGTATAAGGGATCTTAGGTTCGTTAAAAATGGGATTTTTTCCTTTAGAG 842

Qy       2204 TAGACGAGGATCATGTCCTTTTGCGATCCGAAGGCCTTACGGGAAAAGTTTTTGGGATTT 2263
Db        843 TAAAATAAAATCATATCTTTTATGTTACCATAGCCTATCCTTTTAAAATTTTTAGGATTG 902

Qy       2264 GAAGCGATGCGGGCGATATGGTTAACGAAGTTTCGCCGGCCAAAGACCTCATCAAGGATG 2323
Db        903 CACTTTATGCGTGTGATTTCATTTCTAAAATTTTGTATGCCAAATATTTCATCTAACATT 962

Qy       2324 AGCTTCACCTCGAACCCGTATTTCTCGTCTATGTGAACGAAGATCAGTCCTGAGTCCGCC 2383
Db        963 ACCTTGACATAATGCCCTATCTTGTAATCTGTATGCACATAGATAGAGCCTTGTTCTGAA 1022

Qy       2384 ATCAGCTCCCTGAGAAGTATCAAGCGCTCCCTCAGGAACTCCACAAACTGAGGACCATCG 2443
Db       1023 AGCAATTCTTTAAGCAATACCAGGCGTTGTTTTAAAAATTCCATAAAATCCATACCCACT 1082

Qy       2444 AGGGTGTCATCGTAGCCCAACTGACCGTTTTTGGGCTGGCTGACGGTAGCAACGCGATCT 2503
Db       1083 ACTTTATCACTATAAGCAATATCGCCATTCTTAGAATTGCTAATTGTGGTTGCTCTACCA 1142

Qy       2504 GTTTCATCGCCGCCAACGAGAAACTGCTGGCCGGTTCCATAAGGCGGGTCAATATAGACC 2563
Db       1143 TTT---------GTGATAGTAAAATGATTGTTTGTAGCAAATGGAGGGTCAATATATATC 1193

Qy       2564 AACTGGACCTTCCC 2577
Db       1194 AAATCTATTTTTCC 1207


RESULT 9
AF328909
LOCUS       AF328909               4931 bp    DNA     linear   BCT 12-NOV-2002
DEFINITION  Helicobacter pylori strain CR2 putative cytosine-specific DNA
            methyltransferase, putative HP0052-like protein, type II DNA
            modification enzyme, and type II restriction enzyme genes, complete
            cds.
ACCESSION   AF328909
VERSION     AF328909.1  GI:17225495
KEYWORDS    .
SOURCE      Helicobacter pylori
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
            Helicobacteraceae; Helicobacter.
REFERENCE   1  (bases 1 to 4931)
  AUTHORS   Chanto,G., Occhialini,A., Gras,N., Alm,R.A., Megraud,F. and
            Marais,A.
  TITLE     Identification of strain-specific genes located outside the
            plasticity zone in nine clinical isolates of Helicobacter pylori
  JOURNAL   Microbiology 148 (Pt 11), 3671-3680 (2002)
   PUBMED   12427957
REFERENCE   2  (bases 1 to 4931)
  AUTHORS   Chanto,G., Occhialini,A., Gras,N., Megraud,F. and Marais,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-DEC-2000) Laboratoire de Bacteriologie, Universite
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .4931
                     /organism="Helicobacter pylori"
                     /mol_type="genomic DNA"
                     /strain="CR2"
                     /specific_host="Homo sapiens"
                     /db_xref="taxon:210"
                     /map="region F between JHP43/HP50 and JHP47/HP55"
                     /note="isolated from patient with gastric carcinoma"
     CDS             complement(1. .1086)
                     /note="putative HP0051-like protein"
                     /codon_start=1
                     /transl_table=11
                     /product="putative cytosine-specific DNA
                     methyltransferase"
                     /protein_id="AAL37432.1"
                     /db_xref="GI:17225496"
                     /translation="MNYKILDLFCGAGGFSAGLGYLKEFDALIGLDCDKQALITFENN
                     HKNAIGVCGDITQTEIKEKVIELAKKLEINMIIGGPPCQGFSNKGKNLGLKDPRNFLF
                     LEYIEIVKAIKPEIFIIENVKNLISCAKGYFLEEIKERLNALGYQLSYQILNAKDYGV
                     PQNRERAFIVGASRFSFDFNLLEPSQSVNVQDAISDLAYLCSNEGAFESDYLNPIQSS
                     YQALMRKDSPKLYNHQATNHSQAALEKLKLINKEQGKECLPKNLHGKQQFKSTWGRLN
                     WNKISPTIDTRFDTPSNGTNSHPELHRSITPREAARIQSFSDDYIFYGNKTSVCKQIG
                     NAVPPLLALALGKAILKSARNDTNLSR"
     CDS             complement(1083. .2309)
                     /codon_start=1
                     /transl_table=11
                     /product="putative HP0052-like protein"
                     /protein_id="AAL37433.1"
                     /db_xref="GI:17225497"
                     /translation="MLVSRFLNDDIDPFNLGVLLSRFQIKNGCIYGVCSYKASKFVPD
                     YEKSKAQVLNALNTLSVHPIWQSNQESVTKIKGTFVFILEKDLHLDENAFYKKLLNLI
                     IDNDFFNRSHSMTPNQKRFLSGFFESRGSIDTQRNFLTSDYFFHSPLEFNKFHYLIDH
                     FNIPSEALNFNFRELQPEHAQGINQRNAQFRIYLNWYLHHIGLFNPYKARIAEHVFKT
                     TLIYDGIYHKLSYPPTTKYHGNSFTERAHFYLKNVYQQDLDDKSIEKLREQLGWIQKS
                     EEFKRDSKIINFYRISTLNVCSACCGDYDIKERSFISLPLYKITQNPNSYYTEIHHVV
                     SLGKNKELDVLENLAKLCPTCHRALKKGSSEEEFQKRLIRNILNHNKDNLEFAQLRFE
```

```
                TDDFLTLIDKIHESLK"
     CDS        complement(2677. .3708)
                /function="methyltransferase"
                /note="JHP45-like protein"
                /codon_start=1
                /transl_table=11
                /product="type II DNA modification enzyme"
                /protein_id="AAL37434.1"
                /db_xref="GI:17225498"
                /translation="MQNLLIQAENAIALLFLLNDKNLKGKIDLIYIDPPFATNNHFTI
                TNGRATTISNSKNGDIAYSDKVVGMDFIEFLKQRLVLLKELLSEQGSIYVHTDCKIGH
                YVKVMLDEIFGIQNFRNEITRIKCNPKNFKRMGYGNIKDMILFYSKGKNPIFNEPKIP
                YTPQDLEKRFPKIDKDKRRYTTVPIHAPGEVESGECSKAFKGMLPPKGRHWRTDIATL
                ERWDKEGLIEYSNNNNPRKKIYALEQVGKRVQDIWEFKDPQYPSYPTEKNAQLLDLII
                KTSSNKDSIVLDCFCGSGTTLKSAFLLQRKFIGIDNSNLAIQACKNKLETITKDLFVS
                QNFYDFLVF"
     CDS        complement(3718. .4503)
                /note="JHP46-like protein"
                /codon_start=1
                /transl_table=11
                /product="type II restriction enzyme"
                /protein_id="AAL37435.1"
                /db_xref="GI:17225499"
                /translation="MVNYWTKLSIEYANQRSYLDDLFQVYPTIPEGLREIDSKIWSNI
                EYHFKQKDNLALITELLNLDLFPIKDSYMAYLKRDKSALERNPRTINRICGRLYEMGL
                NKIFEKCSEPKETNRQIGPMFKDWLNNKSLGVEPVDLNDFIANENDAILRASDNIMAE
                FAKSHLNYHHHKGLDFVARFNKKYIIGEAKFLTDFGGHQNAQFNDAISTIEAPNIKAI
                KVAILDGVLYIESNNKMRKLLDTTYKNYNIMSALVLREFLYQI"
ORIGIN

  Query Match             3.1%;  Score 184;  DB 1;  Length 4931;
  Best Local Similarity   51.9%;  Pred. No. 1.2e-28;
  Matches  474;  Conservative    0;  Mismatches  425;  Indels   15;  Gaps    2;

Qy       1664 ACCCTTCTAGTGAAGGCTTTGACCGCCTCCCAGGAGGCATCTATGCCGATGGATCGCCGC 1723
Db       2738 AGCTTGTTTTTGCAAGCTTGGATAGCTAAATTGGAATTATCAATGCCTATAAATTTTCGT 2797

Qy       1724 TTTAAGAGGGGTGAGGCTATAAGCGTAGTACCGGAGCCTGCGAAGGGATCGAGCACTAAA 1783
Db       2798 TGCAATAAAAACGCAGATTTTAAGGTTGTTCCAGAACCACAAAAACAATCTAAAACAATA 2857

Qy       1784 TCCCCCTCGTTACTCCCTGTTTGGACGATGAGCTTGAGCATGTCCAGATTTTTCTCGGTG 1843
Db       2858 CTATCTTTATTAGAAGAGGTTTTAATGATTAAGTCTAATAATTGAGCGTTTTTTTCTGTA 2917

Qy       1844 GGGTATCGCGGGTACGGAGGATCCTTGAACTGCCAAACGTCCTGGAGCTTCTTCCCCTTC 1903
Db       2918 GGGTAGCTTGGATATTGTGGGTCTTTAAATTCCCAGATGTCTTGGACTCTTTTGCC---- 2973

Qy       1904 TTCAGGCGATCCCGAGCGTAAACTTTCTTCCGCGGCACCCCGTTCTTTGACCAGACAATA 1963
Db       2974 --AACTTGTTCTAAGGCATAAATTTTTTTTCTAGGGTTATTATTGTTAGAATACTCAATC 3031

Qy       1964 AGCCCTTGAGCGTCTAGCTCGTCAAGCTTCTCCGGGGGGATAGCGCCAATGCCGTCCAGGA 2023
Db       3032 AAACCTTCTTTATCCCAACGCTCAAGTGTGGCAATATCAGTGCGCCAATGCCGCCCTTTT 3091

Qy       2024 GGGGGAAGTATTCCTCGCCAAGGCCTTCCGGTAGGGCCATCCTTGGTTTCTCCAGGAGCA 2083
Db       3092 GGAGGTAGCATACCTTTAAATGCTTTAGAACATTCGCCACTTTCCACTTCTCCTGGAGCA 3151

Qy       2084 TGCAGGGGATTGGTGGTGTACCGTTCCCCGTTCTCGTCTACAAAGGGGAAAAGCCTAGCG 2143
Db       3152 TGTATTGGAACGGTAGTGTAACGCCTTTTGTCTTTGTCAATTTTAGGGAATCGTTTTTCT 3211

Qy       2144 ATCTCCTCTTCCGAATAGGGGCTAGCCGATTCGTTCCAAACGTAGTCCCGCGTTTTGGAG 2203
Db       3212 AAATCTTGTGGCGTATAAGGGATCTTAGGTTCGTTAAAAATGGGATTTTTTCCTTTAGAG 3271

Qy       2204 TAGACGAGGATCATGTCCTTTTGCGATCCGAAGGCCTTACGGGAAAAGTTTTTGGGATTT 2263
Db       3272 TAAAATAAAATCATATCTTTTATGTTGCCATAGCCCATCCTTTTAAAATTTTTAGGATTG 3331

Qy       2264 GAAGCGATGCGGGCGATATGGTTAACGAAGTTTCGCCGGCCAAAGACCTCATCAAGGATG 2323
Db       3332 CACTTTATGCGTGTGATTTCATTTCTAAAATTTTGTATGCCAAATATTTCATCTAACATC 3391

Qy       2324 AGCTTCACCTCGAACCCGTATTTCTCGTCTATGTGAACGAAGATCAGTCCTGAGTCCGCC 2383
Db       3392 ACCTTAACATAATGCCCTATCTTACAATCTGTATGCACATAGATAGAGCCTTGTTCTGAA 3451

Qy       2384 ATCAGCTCCCTGAGAAGTATCAAGCGCTCCCTCAGGAACTCCACAAACTGAGGACCATCG 2443
Db       3452 AGCAATTCTTTAAGCAATACCAGGCGTTGTTTTAAAAATTCTATAAAATCCATACCCACT 3511

Qy       2444 AGGGTGTCATCGTAGCCCAACTGACCGTTTTTGGGCTGGCTGACGGTAGCAACGCGATCT 2503
Db       3512 ACTTTGTCACTATAAGCAATATCGCCATTCTTAGAATTGCTAATTGTGGTTGCTCTGCCA 3571

Qy       2504 GTTTCATCGCCGCCAACGAGAAACTGCTGGCCGGTTCCATAAGGCGGGTCAATATAGACC 2563
Db       3572 TTT---------GTGATAGTAAAATGATTGTTTGTAGCAAATGGAGGGTCAATATATATC 3622

Qy       2564 AACTGGACCTTCCC 2577
Db       3623 AAATCTATTTTTCC 3636


RESULT 10
AF328912
LOCUS       AF328912               3172 bp    DNA     linear   BCT 12-NOV-2002
DEFINITION  Helicobacter pylori strain CR14 JHP44-like protein, type II DNA
            modification enzyme, and type II restriction enzyme genes, complete
            cds.
ACCESSION   AF328912
VERSION     AF328912.1  GI:17225508
KEYWORDS    .
SOURCE      Helicobacter pylori
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
            Helicobacteraceae; Helicobacter.
REFERENCE   1  (bases 1 to 3172)
  AUTHORS   Chanto,G., Occhialini,A., Gras,N., Alm,R.A., Megraud,F. and
            Marais,A.
  TITLE     Identification of strain-specific genes located outside the
            plasticity zone in nine clinical isolates of Helicobacter pylori
  JOURNAL   Microbiology 148 (Pt 11), 3671-3680 (2002)
   PUBMED   12427957
REFERENCE   2  (bases 1 to 3172)
  AUTHORS   Chanto,G., Occhialini,A., Gras,N., Megraud,F. and Marais,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-DEC-2000) Laboratoire de Bacteriologie, Universite
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .3172
                     /organism="Helicobacter pylori"
                     /mol_type="genomic DNA"
                     /strain="CR14"
                     /specific_host="Homo sapiens"
                     /db_xref="taxon:210"
                     /map="region F between JHP43/HP50 and JHP47/HP55"
                     /note="isolated from patient with duodenal ulcer"
     CDS             complement(1. .558)
                     /function="methyltransferase"
                     /note="type II DNA modification enzyme"
                     /codon_start=1
                     /transl_table=11
                     /product="JHP44-like protein"
                     /protein_id="AAL37442.1"
                     /db_xref="GI:17225509"
                     /translation="MLVSRFLNAIDPFNLGVLLSRFQIKNGCIYGVCSYRASKFTSGY
                     EESKARVLNALNTLSAHQIWQSNQESVTKVKGTFVFILENDLHLDENAFYKKLLNLLI
                     DNDFFNRSHLVTPSNGTNSHPELHRSITPREATRIQSFSDDYIFYGNKTSVCKQIGNA
                     VPPLLALALGKAILKSARNDTNLSR"
     CDS             complement(921. .1868)
```

```
                     /function="methyltransferase"
                     /note="JHP45-like protein"
                     /codon_start=1
                     /transl_table=11
                     /product="type II DNA modification enzyme"
                     /protein_id="AAL37443.1"
                     /db_xref="GI:17225510"
                     /translation="MIYIDPPFATNNHFTITNGRATTISNSKNGDIAYSDKVVGMDFI
                     EFLKQRLVLLKELLSEQGSIYVHTDYKIGHYVKVMLDEIFGIQNFRNEITRIKCNPKN
                     FKRIGYGNIKDMILFYSKGKNPIFNEPKIPYTPQDLEKRFPKIDKDKRRYTTVPIHAP
                     GEVESGECSKAFKGMLPPKGRHWRTDIATLERWDKEGLIEYSNNNNPRKKIYALEQVG
                     KRVQDIWEFKDPQYPSYPTEKNAQLLDLIIKTSSNKDSIVLDCFCGSGTTLKSAFLLQ
                     RKFIGIDNSGLAIKACKNKLETITKDLFISQNFYDFLVF"
     CDS             complement(1963. .2652)
                     /note="JHP46-like protein"
                     /codon_start=1
                     /transl_table=11
                     /product="type II restriction enzyme"
                     /protein_id="AAL37444.1"
                     /db_xref="GI:17225511"
                     /translation="MREIDSKIWSNIEYHFKQKDNLTLITELLNLDLFPIKDSYIAYL
                     KRDKSTLERNPRTINCICGRLYEMGVNKIFEKCSEPKETNRQIGPMFKDWLNNKSLGV
                     EPVDLNDFIANENDAILKASDNVMAEFAKSHLNYHHHKGLDFVARFNKKYIIGEAKFL
                     TDFGGHQNAQFNDAISTIEAPNIKAIKVAILDGVLYIESNNKMRKLLDTTYKNYHIMS
                     ALVLREFLYQI"
ORIGIN

  Query Match             3.1%;  Score 182.4;  DB 1;  Length 3172;
  Best Local Similarity   51.8%;  Pred. No. 2.8e-28;
  Matches  473;  Conservative    0;  Mismatches  426;  Indels   15;  Gaps    2;

Qy       1664 ACCCTTCTAGTGAAGGCTTTGACCGCCTCCCAGGAGGCATCTATGCCGATGGATCGCCGC 1723
Db        982 AGCTTGTTTTTGCAAGCTTTGATAGCTAAACCGGAATTATCAATGCCTATAAATTTTCGT 1041

Qy       1724 TTTAAGAGGGGTGAGGCTATAAGCGTAGTACCGGAGCCTGCGAAGGGATCGAGCACTAAA 1783
Db       1042 TGCAATAAAAACGCAGATTTTAAGGTTGTTCCAGAACCACAAAAACAATCTAAAACAATG 1101

Qy       1784 TCCCCCTCGTTACTCCCTGTTTGGACGATGAGCTTGAGCATGTCCAGATTTTTCTCGGTG 1843
Db       1102 CTATCTTTATTAGAAGAGGTTTTAATGATTAAATCTAATAATTGAGCGTTTTTTTCTGTA 1161

Qy       1844 GGGTATCGCGGGTACGGAGGATCCTTGAACTGCCAAACGTCCTGGAGCTTCTTCCCCTTC 1903
Db       1162 GGGTAGCTTGGATATTGTGGGTCTTTAAATTCCCAGATGTCTTGGACTCTTTTGCC---- 1217

Qy       1904 TTCAGGCGATCCCGAGCGTAAACTTTCTTCCGCGGCACCCCGTTCTTTGACCAGACAATA 1963
Db       1218 --AACTTGTTCTAAGGCATAAATTTTTTTTCTAGGGTTATTATTGTTAGAATACTCAATC 1275

Qy       1964 AGCCCTTGAGCGTCTAGCTCGTCAAGCTTCTCCGGGGGATAGCGCCAATGCCGTCCAGGA 2023
Db       1276 AAACCTTCTTTATCCCAACGCTCAAGTGTGGCAATATCAGTGCGCCAATGCCGCCCTTTT 1335

Qy       2024 GGGGGAAGTATTCCTCGCCAAGGCCTTCCGGTAGGGCCATCCTTGGTTTCTCCAGGAGCA 2083
Db       1336 GGAGGTAGCATACCTTTAAATGCTTTAGAACATTCGCCACTTTCTACTTCTCCTGGAGCA 1395

Qy       2084 TGCAGGGGATTGGTGGTGTACCGTTCCCCGTTCTCGTCTACAAAGGGGAAAAGCCTAGCG 2143
Db       1396 TGTATTGGAACGGTAGTGTAACGCCTTTTGTCTTTGTCAATTTTAGGGAATCGTTTTTCT 1455

Qy       2144 ATCTCCTCTTCCGAATAGGGGCTAGCCGATTCGTTCCAAACGTAGTCCCGCGTTTTGGAG 2203
Db       1456 AAATCTTGTGGCGTATAAGGGATCTTAGGTTCGTTAAAAATGGGATTTTTTCCTTTAGAG 1515

Qy       2204 TAGACGAGGATCATGTCCTTTTGCGATCCGAAGGCCTTACGGGAAAAGTTTTTGGGATTT 2263
Db       1516 TAAAATAAAATCATATCTTTTATATTACCATAGCCTATCCTTTTAAAATTTTTAGGATTG 1575

Qy       2264 GAAGCGATGCGGGCGATATGGTTAACGAAGTTTCGCCGGCCAAAGACCTCATCAAGGATG 2323
Db       1576 CACTTTATGCGTGTGATTTCATTTCTAAAATTTTGTATGCCAAATATTTCATCTAACATT 1635

Qy       2324 AGCTTCACCTCGAACCCGTATTTCTCGTCTATGTGAACGAAGATCAGTCCTGAGTCCGCC 2383
Db       1636 ACCTTAACATAATGTCCTATCTTATAATCTGTATGCACATAGATAGAGCCTTGTTCTGAA 1695

Qy       2384 ATCAGCTCCCTGAGAAGTATCAAGCGCTCCCTCAGGAACTCCACAAACTGAGGACCATCG 2443
Db       1696 AGCAATTCTTTAAGCAATACTAGGCGTTGCTTTAAAAATTCTATAAAATCCATACCCACT 1755

Qy       2444 AGGGTGTCATCGTAGCCCAACTGACCGTTTTTGGGCTGGCTGACGGTAGCAACGCGATCT 2503
Db       1756 ACTTTGTCGCTATAAGCAATATCGCCATTCTTAGAATTGCTAATTGTGGTTGCTCTGCCA 1815

Qy       2504 GTTTCATCGCCGCCAACGAGAAACTGCTGGCCGGTTCCATAAGGCGGGTCAATATAGACC 2563
Db       1816 TTT---------GTGATAGTAAAATGATTGTTTGTAGCAAATGGAGGGTCAATATATATC 1866

Qy       2564 AACTGGACCTTCCC 2577
Db       1867 AAATCTATTTTTCC 1880
```

```
RESULT 11
AF328911
LOCUS       AF328911                3261 bp    DNA     linear   BCT 12-NOV-2002
DEFINITION  Helicobacter pylori strain CR9 type II DNA modification enzyme,
            type II DNA modification enzyme, and type II restriction enzyme
            genes, complete cds.
ACCESSION   AF328911
VERSION     AF328911.1  GI:17225504
KEYWORDS    .
SOURCE      Helicobacter pylori
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
            Helicobacteraceae; Helicobacter.
REFERENCE   1  (bases 1 to 3261)
  AUTHORS   Chanto,G., Occhialini,A., Gras,N., Alm,R.A., Megraud,F. and
            Marais,A.
  TITLE     Identification of strain-specific genes located outside the
            plasticity zone in nine clinical isolates of Helicobacter pylori
  JOURNAL   Microbiology 148 (Pt 11), 3671-3680 (2002)
   PUBMED   12427957
REFERENCE   2  (bases 1 to 3261)
  AUTHORS   Chanto,G., Occhialini,A., Gras,N., Megraud,F. and Marais,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-DEC-2000) Laboratoire de Bacteriologie, Universite
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .3261
                     /organism="Helicobacter pylori"
                     /mol_type="genomic DNA"
                     /strain="CR9"
                     /specific_host="Homo sapiens"
                     /db_xref="taxon:210"
                     /map="region F between JHP43/HP50 and JHP47/HP55"
                     /note="isolated from patient with gastric carcinoma"
     CDS             complement(1. .558)
                     /function="methyltransferase"
                     /note="JHP44-like protein"
                     /codon_start=1
                     /transl_table=11
                     /product="type II DNA modification enzyme"
                     /protein_id="AAL37439.1"
                     /db_xref="GI:17225505"
                     /translation="MLVSRFLNAIDPFNLGVLLSRFQIKNGCIYGVCSYKASKFTSSY
                     EESKAQVLNALNILSKHQIWQSNQESVTKVKGTFVFILENDLNLDENAFYKKLLNLLI
                     DNDFFNRPHLVTPSNGTNSHPELHRSITPREAARIQSFSDNYIFYGNKTSVCKQIGNA
                     VPPLLALALGKAILKSTRNDTNLSR"
     CDS             complement(1008. .2039)
                     /function="methyltransferase"
                     /note="JHP45-like protein"
                     /codon_start=1
```

```
                     /transl_table=11
                     /product="type II DNA modification enzyme"
                     /protein_id="AAL37440.1"
                     /db_xref="GI:17225506"
                     /translation="MQNLLIQAENSIALLFLLNDKNLKGKIDLIYIDPPFATNNHFTI
                     TNGRATTISNSKNGDIAYSDKVVGMDFIEFLKQRLVLLKELLSEQGSIYVHTDYKIGH
                     YVKVMLDEIFGIQNFRNEITRIKCNPKNFKRIGYGNIKDMILFYSKGKNPIFNEPKIP
                     YTPQDLEKRFPKIDKDKRRYTTVPIHAPGEVESGECSKAFKGMLPPKGRHWRTDIATL
                     ERWDKEGLIEYSNNNNPRKKIYALEQAGKRVQDIWEFKDPQYPSYPTEKNAQLLDLII
                     KTSSNKDSIVLDCFCGSGTTLKSAFLLQRKFIGIDNSDLAIQACKNKLETITKDLFVS
                     QNFYDFLVF"
     CDS             complement(2049. .2834)
                     /note="JHP46-like protein"
                     /codon_start=1
                     /transl_table=11
                     /product="type II restriction enzyme"
                     /protein_id="AAL37441.1"
                     /db_xref="GI:17225507"
                     /translation="MVNYWTKLSIEYANQRSYLDDLFQVYPTIPEGLREIDSKIWSNI
                     EYHFKQKDNLALITELLNLDLFPIKDSYIAYLKRDKSTLERNPRTINRICGRLYEMGL
                     NKIFEKCSEPKETNRQIGPMFKDWLNNKSLGVEPVDLNDFIANENDAILKASDNVMAE
                     FAKSHLNYHHHKGLDFVARFNKKYIIGEAKFLTDFGGHQNAQFNDAISTIKAPNIKAI
                     KVAILDGVLYIESNNKMRKLLDTTYRNYHIMSALVLREFLYQI"
ORIGIN

  Query Match             3.1%;  Score 180.8;  DB 1;  Length 3261;
  Best Local Similarity   51.6%;  Pred. No. 6.1e-28;
  Matches  472;  Conservative    0;  Mismatches  427;  Indels   15;  Gaps    2;

Qy       1664 ACCCTTCTAGTGAAGGCTTTGACCGCCTCCCAGGAGGCATCTATGCCGATGGATCGCCGC 1723
Db       1069 AGCTTGTTTTGCAAGCTTGGATAGCCAAATCGGAATTATCAATGCCTATAAATTTTCGT 1128

Qy       1724 TTTAAGAGGGGTGAGGCTATAAGCGTAGTACCGGAGCCTGCGAAGGGATCGAGCACTAAA 1783
Db       1129 TGCAATAAAAACGCAGATTTTAAAGTTGTTCCAGAACCACAAAAACAATCTAAAACAATA 1188

Qy       1784 TCCCCCTCGTTACTCCCTGTTTGGACGATGAGCTTGAGCATGTCCAGATTTTTCTCGGTG 1843
Db       1189 CTATCTTTATTAGAAGAGGTTTTAATGATTAAGTCTAATAATTGAGCGTTTTTTCTGTG 1248

Qy       1844 GGGTATCGCGGGTACGGAGGATCCTTGAACTGCCAAACGTCCTGGAGCTTCTTCCCCTTC 1903
Db       1249 GGGTAGCTTGGATATTGTGGGTCTTTAAATTCCCAGATGTCTTGGACTCTTTTGCCAGCT 1308

Qy       1904 TTCAGGCGATCCCGAGCGTAAACTTTCTTCCGCGGCACCCCGTTCTTTGACCAGACAATA 1963
Db       1309 T-----GTTCTAAGGCATAAATTTTTTTTCTAGGGTTATTATTGTTAGAATACTCAATC 1362

Qy       1964 AGCCCTTGAGCGTCTAGCTCGTCAAGCTTCTCCGGGGGATAGCGCCAATGCCGTCCAGGA 2023
Db       1363 AAACCTTCTTTATCCCAACGCTCAAGTGTGGCAATATCAGTGCGCCAATGCCGCCCTTTT 1422

Qy       2024 GGGGGAAGTATTCCTCGCCAAGGCCTTCCGGTAGGGCCATCCTTGGTTTCTCCAGGAGCA 2083
Db       1423 GGAGGTAGCATACCTTTAAATGCTTTAGAGCATTCGCCACTTTCTACTTCTCCTGGAGCA 1482

Qy       2084 TGCAGGGGATTGGTGGTGTACCGTTCCCCGTTCTCGTCTACAAAGGGGAAAAGCCTAGCG 2143
Db       1483 TGTATTGGAACGGTAGTGTAACGCCTTTTATCCTTGTCAATTTTAGGGAATCGTTTTTCT 1542

Qy       2144 ATCTCCTCTTCCGAATAGGGGCTAGCCGATTCGTTCCAAACGTAGTCCCGCGTTTTGGAG 2203
Db       1543 AAATCTTGTGGCGTATAAGGAATCTTAGGCTCGTTAAAAATGGATTTTTTCCTTTAGAA 1602

Qy       2204 TAGACGAGGATCATGTCCTTTTGCGATCCGAAGGCCTTACGGGAAAAGTTTTTGGGATTT 2263
Db       1603 TAAAATAAAATCATATCTTTTATGTTACCATAGCCTATCCTTTTAAAATTTTTAGGATTG 1662

Qy       2264 GAAGCGATGCGGGCGATATGGTTAACGAAGTTTCGCCGGCCAAAGACCTCATCAAGGATG 2323
Db       1663 CACTTTATACGTGTGATTTCATTTCTAAAATTTTGTATGCCAAATATTTCATCTAACATT 1722

Qy       2324 AGCTTCACCTCGAACCCGTATTTCTCGTCTATGTGAACGAAGATCAGTCCTGAGTCCGCC 2383
Db       1723 ACCTTAACATAATGTCCTATCTTATAATCTGTATGCACATAGATAGAGCCTTGTTCTGAA 1782

Qy       2384 ATCAGCTCCCTGAGAAGTATCAAGCGCTCCCTCAGGAACTCCACAAACTGAGGACCATCG 2443
Db       1783 AGCAATTCTTTAAGCAATACTAGGCGTTGCTTTAAAAATTCTATAAAATCCATACCCACT 1842

Qy       2444 AGGGTGTCATCGTAGCCCAACTGACCGTTTTTGGGCTGGCTGACGGTAGCAACGCGATCT 2503
Db       1843 ACTTTGTCGCTATAAGCAATATCGCCATTCTTAGAATTGCTAATTGTGGTTGCTCTACCA 1902

Qy       2504 GTTTCATCGCCGCCAACGAGAAACTGCTGGCCGGTTCCATAAGGCGGGTCAATATAGACC 2563
Db       1903 TTT---------GTGATAGTAAAATGATTGTTTGTAGCAAATGGAGGGTCAATATATATC 1953

Qy       2564 AACTGGACCTTCCC 2577
Db       1954 AAATCTATTTTTCC 1967


RESULT 12
AF328915
LOCUS       AF328915                5001 bp    DNA     linear   BCT 12-NOV-2002
DEFINITION  Helicobacter pylori strain CR29 cytosine-specific DNA
            methyltransferase, putative HP0052-like protein, type II DNA
            modification enzyme, and type II restriction enzyme genes, complete
            cds.
ACCESSION   AF328915
VERSION     AF328915.1  GI:17225516
KEYWORDS    .
SOURCE      Helicobacter pylori
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
            Helicobacteraceae; Helicobacter.
REFERENCE   1  (bases 1 to 5001)
  AUTHORS   Chanto,G., Occhialini,A., Gras,N., Alm,R.A., Megraud,F. and
            Marais,A.
  TITLE     Identification of strain-specific genes located outside the
            plasticity zone in nine clinical isolates of Helicobacter pylori
  JOURNAL   Microbiology 148 (Pt 11), 3671-3680 (2002)
   PUBMED   12427957
REFERENCE   2  (bases 1 to 5001)
  AUTHORS   Chanto,G., Occhialini,A., Gras,N., Megraud,F. and Marais,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-DEC-2000) Laboratoire de Bacteriologie, Universite
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .5001
                     /organism="Helicobacter pylori"
                     /mol_type="genomic DNA"
                     /strain="CR29"
                     /specific_host="Homo sapiens"
                     /db_xref="taxon:210"
                     /map="region F between JHP43/HP50 and JHP47/HP55"
                     /note="isolated from patient with chronic gastritis"
     CDS             complement(1. .1086)
                     /note="HP0051-like protein"
                     /codon_start=1
                     /transl_table=11
                     /product="cytosine-specific DNA methyltransferase"
                     /protein_id="AAL37447.1"
                     /db_xref="GI:17225517"
                     /translation="MNYKILDLFCGAGGFSAGLECLKEFDALIGLDCDKQALITFENN
                     HKNATGICGDITQIEIKEKVIKLAQTLEINMIIGGPPCQGFSNKGKNLGLKDPRNFLF
                     LEYIEIVKAIKPEIFIIENVKNLISCAKGYFLEEIKERLNALGYQLSYQILNAKDYGV
                     PQNRERAFIVGASRFSFDFNLLEPSQSVNIQDAISDLAYLCSNEGAFESDYLNPIQSS
                     YQALMRKDSPKLYNHQATNHSQAALEKLKLINKEQGKECLPKNLHGKQQFKSTWGRLN
                     WNKISPTIDTRFDTPSNGTNSHPELHRSITPREATRIQSFSDDYIFYGNKTSVCKQIG
                     NAVPPLLALALGKAILKSARNDTNLSR"
     CDS             complement(1083. .2312)
                     /codon_start=1
                     /transl_table=11
```

```
                     /product="putative HP0052-like protein"
                     /protein_id="AAL37448.1"
                     /db_xref="GI:17225518"
                     /translation="MLVSRFLNAIDPFNLGVLLSRFQIKNGCIYGVCSYKLSKFIPSY
                     EESKARVLNALNTLSKHQIWQSNQESVTKIKGTFVFILENDLHLDENAFYKKLLNLLI
                     DNDFFNRSHLINSMTPNQKRFLSGFFESRGSIDTQRNFLTSDHFFHSPLEFNKFHYLI
                     DHFTIPSEALNFNFRELQPEYVQGIKRNAQFRIYLDWYLYHISLFNPYKAKIAEHVFK
                     TTLVNDNGYYKLSYPPRTKYHGNSFIERAHFYLKNVYQQDLDDKSIEKLREQLGFIQD
                     SEEFRRDSKIISFYRLSTPNVCNACYDDYDIKERSFISLPLYQITQRSDSYYTEIHHV
                     ISLGKDKELDVLENLAKLCPTCHRALRKGASAEEFQKRLIRKILKRNKDNLEFAKLRF
                     ETDDFLTLIDKVYESLK"
     CDS             complement(2752..3783)
                     /function="methyltransferase"
                     /note="JHP45-like protein"
                     /codon_start=1
                     /transl_table=11
                     /product="type II DNA modification enzyme"
                     /protein_id="AAL37449.1"
                     /db_xref="GI:17225519"
                     /translation="MQNLLIQAENAIALLFLLNDKNLKGKIDLIYIDPPFATNNHFTI
                     TNGRATTISNSKNGDIAYSDKVVGMDFIEFLKQRLVLLKELLSEQGSIYVHTDCKIGH
                     YVKVMLDEIFGIQNFRNEITRIKCNPKNFKRMGYGNIKDMILFYSKGKNPIFNEPKIP
                     YTPQDLEKRFPKIDKDKRRYTTVPIHAPGEVESGECSKAFKGMLPPKGRHWRTDIATL
                     ERWDKEGLIEYSNNNNPRKKIYALEQVGKRVQDIWEFKDPQYPSYPTEKNAQLLDLII
                     KTSSNKNSIVLDCFCGSGTTLKSAFLLQRKFIGIDNSDLAIKACKNKLETITKDLFVS
                     QNFYDFLVF"
     CDS             complement(3793..4578)
                     /note="JHP46-like protein"
                     /codon_start=1
                     /transl_table=11
                     /product="type II restriction enzyme"
                     /protein_id="AAL37450.1"
                     /db_xref="GI:17225520"
                     /translation="MVDYWTKLSIEYANQRSYLDDLFQVYPTIPEGLREIDSKIWSNV
                     EYHFKRKDNLALITELLNLDLFPIKDSYMAYLKRDKSALERNPRTINRICGRLYEMGL
                     NKIFEKCSEPKETNRQIGPMFKDWINNKSLGVEPVDLNDFIANENDAILRASDNIMAE
                     FAKSHLNYHHHKGLDFVARFNKKYIIGEAKFLTDFGGHQNAQFNDAISTIEAPNIKAI
                     KVAILDGVLYIESNNKMRKLLDTTYRNYNIMSALVLREFLYQI"
ORIGIN

  Query Match             3.1%;  Score 180.8;  DB 1;  Length 5001;
  Best Local Similarity   51.6%;  Pred. No. 6e-28;
  Matches  472;  Conservative    0;  Mismatches  427;  Indels   15;  Gaps    2;

Qy       1664 ACCCTTCTAGTGAAGGCTTTGACCGCCTCCCAGGAGGCATCTATGCCGATGGATCGCCGC 1723
Db       2813 AGCTTGTTTTTGCAAGCTTTGATAGCCAAATCGGAATTATCAATGCCTATAAATTTTCGT 2872

Qy       1724 TTTAAGAGGGGTGAGGCTATAAGCGTAGTACCGGAGCCTGCGAAGGGATCGAGCACTAAA 1783
Db       2873 TGCAATAAAAACGCAGATTTAAGGTTGTTCCAGAACCACAAAAACAATCTAAAACAATG 2932

Qy       1784 TCCCCCTCGTTACTCCCTGTTTGGACGATGAGCTTGAGCATGTCCAGATTTTTCTCGGTG 1843
Db       2933 CTATTTTTATTAGAAGAGGTCTTAATGATTAAATCTAATAATTGAGCGTTTTTTTCTGTA 2992

Qy       1844 GGGTATCGCGGGTACGGAGGATCCTTGAACTGCCAAACGTCCTGGAGCTTCTTCCCCTTC 1903
Db       2993 GGGTAGCTTGGATATTGTGGGTCTTTAAATTCCCAGATGTCTTGGACTCTTTTGCC---- 3048

Qy       1904 TTCAGGCGATCCCGAGCGTAAACTTTCTTCCGCGGCACCCCGTTCTTTGACCAGACAATA 1963
Db       3049 --AACTTGTTCTAAGGCATAAATTTTTTTTCTAGGGTTATTATTATTAGAGTACTCAATC 3106

Qy       1964 AGCCCTTGAGCGTCTAGCTCGTCAAGCTTCTCCGGGGGATAGCGCCAATGCCGTCCAGGA 2023
Db       3107 AAACCTTCTTTGTCCCAACGCTCAAGTGTGGCAATATCAGTGCGCCAATGCCGCCCTTTT 3166

Qy       2024 GGGGGAAGTATTCCTCGCCAAGGCCTTCCGGTAGGGCCATCCTTGGTTTCTCCAGGAGCA 2083
Db       3167 GGAGGTAGCATACCTTTAAATGCTTTAGAACATTCGCCACTTTCCACTTCTCCTGGAGCA 3226

Qy       2084 TGCAGGGGATTGGTGGTGTACCGTTCCCCGTTCTCGTCTACAAAGGGGAAAAGCCTAGCG 2143
Db       3227 TGTATTGGAACGGTAGTGTAACGCCTTTTATCTTTGTCAATTTTAGGGAATCGTTTTTCT 3286

Qy       2144 ATCTCCTCTTCCGAATAGGGGCTAGCCGATTCGTTCCAAACGTAGTCCCGCGTTTGGAG 2203
Db       3287 AAATCTTGTGGCGTATAAGGAATCTTAGGTTCGTTAAAAATGGGATTTTTCCTTTAGAG 3346

Qy       2204 TAGACGAGGATCATGTCCTTTTGCGATCCGAAGGCCTTACGGGAAAAGTTTTTGGGATTT 2263
Db       3347 TAAAATAAAATCATATCTTTTATGTTGCCATAGCCCATCCTTTTAAAATTTTTAGGATTG 3406

Qy       2264 GAAGCGATGCGGGCGATATGGTTAACGAAGTTTCGCCGGCCAAAGACCTCATCAAGGATG 2323
Db       3407 CACTTTATGCGTGTGATTTCATTTCTAAAATTTTGTATGCCAAATATTTCATCTAACATC 3466

Qy       2324 AGCTTCACCTCGAACCCGTATTTCTCGTCTATGTGAACGAAGATCAGTCCTGAGTCCGCC 2383
Db       3467 ACCTTAACATAATGCCCTATCTTACAATCTGTATGCACATAGATAGAGCCTTGTTCTGAA 3526

Qy       2384 ATCAGCTCCCTGAGAAGTATCAAGCGCTCCCTCAGGAACTCCACAAACTGAGGACCATCG 2443
Db       3527 AGCAATTCTTTAAGCAATACTAGGCGTTGCTTTAAAAATTCTATAAAATCCATACCCACT 3586

Qy       2444 AGGGTGTCATCGTAGCCCAACTGACCGTTTTTGGGCTGGCTGACGGTAGCAACGCGATCT 2503
Db       3587 ACTTTGTCGCTATAAGCAATATCGCCATTCTTAGAATTGCTAATTGTGGTTGCTCTACCA 3646

Qy       2504 GTTTCATCGCCGCCAACGAGAAACTGCTGGCCGGTTCCATAAGGCGGGTCAATATAGACC 2563
Db       3647 TTT---------GTGATAGTAAAATGGTTGTTTGTGGCAAATGGAGGGTCAATATATATC 3697

Qy       2564 AACTGGACCTTCCC 2577
Db       3698 AAATCTATTTTTCC 3711

RESULT 13
AR139425/c
LOCUS       AR139425                 180 bp    DNA     linear   PAT 16-JUN-2001
DEFINITION  Sequence 5 from patent US 6207377.
ACCESSION   AR139425
VERSION     AR139425.1  GI:14481921
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 180)
  AUTHORS   Wayne,J. and Xu,S.-y.
  TITLE     Method for construction of thermus-E. coli shuttle vectors and
            identification of two Thermus plasmid replication origins
  JOURNAL   Patent: US 6207377-A 5 27-MAR-2001;
FEATURES             Location/Qualifiers
     source          1..180
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             3.1%;  Score 180;  DB 6;  Length 180;
  Best Local Similarity   100.0%;  Pred. No. 1e-27;
  Matches  180;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4128 CTCTTCAAAGAAGGTTTTTTCGTTCTTCACCCTCGGACCTCCTTGTCATCTGGAGCCCGA 4187
Db        180 CTCTTCAAAGAAGGTTTTTTCGTTCTTCACCCTCGGACCTCCTTGTCATCTGGAGCCCGA 121

Qy       4188 GGCGTTACCCTAGGTCTTGGGGGTGATCCGGGGCAACCGCCTCGGTTTCGCCTTTTTATG 4247
Db        120 GGCGTTACCCTAGGTCTTGGGGGTGATCCGGGGCAACCGCCTCGGTTTCGCCTTTTTATG 61

Qy       4248 GGTCCAAAATAACCGTCAGCCCAGCGGCTGGCAATCCCCCCTCCTAAAAGGCCGTTATAG 4307
Db         60 GGTCCAAAATAACCGTCAGCCCAGCGGCTGGCAATCCCCCCTCCTAAAAGGCCGTTATAG 1
```

RESULT 14
AF328910
LOCUS AF328910 3184 bp DNA linear BCT 12-NOV-2002
DEFINITION Helicobacter pylori strain CR4 type II DNA modification enzyme, type II DNA modification enzyme, and type II restriction enzyme genes, complete cds.
ACCESSION AF328910
VERSION AF328910.1 GI:17225500
KEYWORDS .
SOURCE Helicobacter pylori
ORGANISM Helicobacter pylori
Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales; Helicobacteraceae; Helicobacter.
REFERENCE 1 (bases 1 to 3184)
AUTHORS Chanto,G., Occhialini,A., Gras,N., Alm,R.A., Megraud,F. and Marais,A.
TITLE Identification of strain-specific genes located outside the plasticity zone in nine clinical isolates of Helicobacter pylori
JOURNAL Microbiology 148 (Pt 11), 3671-3680 (2002)
PUBMED 12427957
REFERENCE 2 (bases 1 to 3184)
AUTHORS Chanto,G., Occhialini,A., Gras,N., Megraud,F. and Marais,A.
TITLE Direct Submission
JOURNAL Submitted (12-DEC-2000) Laboratoire de Bacteriologie, Universite Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES Location/Qualifiers
source 1. .3184
/organism="Helicobacter pylori"
/mol_type="genomic DNA"
/strain="CR4"
/specific_host="Homo sapiens"
/db_xref="taxon:210"
/map="region F between JHP43/HP50 and JHP47/HP55"
/note="isolated from patient with gastric carcinoma"
CDS complement(1. .561)
/function="methyltransferase"
/note="JHP44-like protein"
/codon_start=1
/transl_table=11
/product="type II DNA modification enzyme"
/protein_id="AAL37436.1"
/db_xref="GI:17225501"
/translation="MLVSRFLNDDIDPFNLGVLLSRFQIKNGCIYGVCSYKASKFTSS
YEESKAQVLNALNILSKHQIWQSNQESVTKVKGTFVFILENDLNLDENAFYKKLLNLL
IDNDFFNRPHLVTPSNGTNSHPELHRSITPREAARIQSFSDNYIFYGNKTSVCKQIGN
AVPPLLALALGKAILKSTRNDTNLSR"
CDS complement(932. .1963)
/function="methyltransferase"
/note="JHP45-like protein"
/codon_start=1
/transl_table=11
/product="type II DNA modification enzyme"
/protein_id="AAL37437.1"
/db_xref="GI:17225502"
/translation="MQNLLIQAENAIALLFLLNDKNLKGKIDLIYIDPPFATNNHFTI
TNGRATTISNSKNGDIAYSDKVVGMDFIEFLKQRLVLLKELLSEQGSIYVHTDCKIGH
YVKVMLDEIFGIQNFRNEITRIKCNPKNFKRIGYGNIKDMILFYSKGKNPIFNEPKIP
YTPQDLEKRFPKIDKDKRRYTTVPIHAPGEVESGECSKAFKGMLPPKGRHWRTDIATL
ERWDKEGLIEYSNNNNPRKKIYALEQAGKRVQDIWEFKDPQYPSYPTEKNAQLLDLII
KTSSNKDSIVLDCFCGSGTTLKSAFLLQRKFIGIDNSGLAIQACKNKLETITKDLFVS
QNFYDFLVF"
CDS complement(1973. .2758)
/note="JHP46-like protein"
/codon_start=1
/transl_table=11
/product="type II restriction enzyme"
/protein_id="AAL37438.1"
/db_xref="GI:17225503"
/translation="MVNYWTKLSIEYANQRSYLDDLFQVYPTIPEGLREIDSKIWSNV
EYHFKRKDNLALITELLNLDLFPIKDSYMAYLKRDKSALERNPRTINRICGRLYEMGL
NKIFEKCSEPKETNRQIGPMFKDWINNKSLGVEPVDLNDFIANENDAILRASDNIMAE
FAKSHLNYHHHKGLDFVARFNKKYIIGEAKFLTDFGGHQNAQFNDAISTIEAPNIKAI
KVAILDGVLYIESNNKMRKLLDTTYRNYNIMSALVLREFLYQI"
ORIGIN

Query Match 3.1%; Score 179.2; DB 1; Length 3184;
Best Local Similarity 51.5%; Pred. No. 1.4e-27;
Matches 471; Conservative 0; Mismatches 428; Indels 15; Gaps 2;

```
Qy       1664 ACCCTTCTAGTGAAGGCTTTGACCGCCTCCCAGGAGGCATCTATGCCGATGGATCGCCGC 1723
Db        993 AGCTTGTTTTTGCAAGCTTGGATAGCTAAACCGGAATTATCAATGCCTATAAATTTTCGT 1052

Qy       1724 TTTAAGAGGGGTGAGGCTATAAGCGTAGTACCGGAGCCTGCGAAGGGATCGAGCACTAAA 1783
Db       1053 TGCAATAAAAATGCAGATTTTAAAGTTGTTCCAGAACCACAAAAACAATCTAAAACAATG 1112

Qy       1784 TCCCCCTCGTTACTCCCTGTTTGGACGATGAGCTTGAGCATGTCCAGATTTTTCTCGGTG 1843
Db       1113 CTATCTTTATTAGAAGAGGTTTTAATGATTAAATCTAATAATTGAGCGTTTTTTTCTGTG 1172

Qy       1844 GGGTATCGCGGGTACGGAGGATCCTTGAACTGCCAAACGTCCTGGAGCTTCTTCCCCTTC 1903
Db       1173 GGGTAGCTTGGATATTGTGGGTCTTTAAATTCCCAGATGTCTTGGACTCTTTTGCCAGCT 1232

Qy       1904 TTCAGGCGATCCCGAGCGTAAACTTTCTTCCGCGGCACCCCGTTCTTTGACCAGACAATA 1963
Db       1233 T------GTTCTAAGGCATAAATTTTTTTTCTAGGGTTATTATTGTTAGAATACTCAATC 1286

Qy       1964 AGCCCTTGAGCGTCTAGCTCGTCAAGCTTCTCCGGGGGATAGCGCCAATGCCGTCCAGGA 2023
Db       1287 AAACCTTCTTTATCCCAACGCTCAAGTGTGGCAATATCAGTGCGCCAATGCCGCCCTTTT 1346

Qy       2024 GGGGGAAGTATTCCTCGCCAAGGCCTTCCGGTAGGGCCATCCTTGGTTTCTCCAGGAGCA 2083
Db       1347 GGAGGTAGCATACCTTTAAATGCTTTAGAGCATTCGCCACTTTCTACTTCTCCTGGAGCA 1406

Qy       2084 TGCAGGGGATTGGTGGTGTACCGTTCCCCGTTCTCGTCTACAAAGGGGAAAAGCCTAGCG 2143
Db       1407 TGTATTGGAACGGTAGTGTAACGCCTTTTATCCTTGTCAATTTTAGGGAATCGTTTTTCT 1466

Qy       2144 ATCTCCTCTTCCGAATAGGGGCTAGCCGATTCGTTCCAAACGTAGTCCCGCGTTTTGGAG 2203
Db       1467 AAATCTTGTGGCGTATAAGGAATCTTAGGCTCGTTAAAAATGGGATTTTTTCCTTTAGAA 1526

Qy       2204 TAGACGAGGATCATGTCCTTTTGCGATCCGAAGGCCTTACGGGAAAAGTTTTTGGGATTT 2263
Db       1527 TAAAATAAAATCATATCTTTTATGTTACCATAGCCTATCCTTTTAAAATTTTTAGGATTG 1586

Qy       2264 GAAGCGATGCGGGCGATATGGTTAACGAAGTTTCGCCGGCCAAAGACCTCATCAAGGATG 2323
Db       1587 CACTTTATACGTGTGATTTCATTTCTAAAATTTTGTATGCCAAATATTTCATCTAACATC 1646

Qy       2324 AGCTTCACCTCGAACCCGTATTTCTCGTCTATGTGAACGAAGATCAGTCCTGAGTCCGCC 2383
Db       1647 ACCTTAACATAATGCCCTATCTTACAATCTGTATGCACATAGATAGAGCCTTGTTCTGAA 1706

Qy       2384 ATCAGCTCCCTGAGAAGTATCAAGCGCTCCCTCAGGAACTCCACAAACTGAGGACCATCG 2443
Db       1707 AGCAATTCTTTAAGCAATACTAGGCGTTGCTTTAAAAATTCTATAAAATCCATACCCACT 1766

Qy       2444 AGGGTGTCATCGTAGCCCAACTGACCGTTTTTGGGCTGGCTGACGGTAGCAACGCGATCT 2503
Db       1767 ACTTTGTCGCTATAAGCAATATCGCCATTCTTAGAATTGCTAATTGTGGTTGCTCTGCCA 1826

Qy       2504 GTTTCATCGCCGCCAACGAGAAACTGCTGGCCGGTTCCATAAGGCGGGTCAATATAGACC 2563
Db       1827 TTT---------GTGATAGTAAAATGATTGTTTGTAGCAAATGGAGGGTCAATATATATC 1877

Qy       2564 AACTGGACCTTCCC 2577
Db       1878 AAATCTATTTTTCC 1891
```

RESULT 15

```
AF328916
LOCUS       AF328916                4931 bp    DNA     linear   BCT 12-NOV-2002
DEFINITION  Helicobacter pylori strain CR35 cytosine-specific DNA
            methyltransferase, putative HP0052-like protein, type II DNA
            modification enzyme, and type II restriction enzyme genes, complete
            cds.
ACCESSION   AF328916
VERSION     AF328916.1  GI:17225521
KEYWORDS    .
SOURCE      Helicobacter pylori
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
            Helicobacteraceae; Helicobacter.
REFERENCE   1  (bases 1 to 4931)
  AUTHORS   Chanto,G., Occhialini,A., Gras,N., Alm,R.A., Megraud,F. and
            Marais,A.
  TITLE     Identification of strain-specific genes located outside the
            plasticity zone in nine clinical isolates of Helicobacter pylori
  JOURNAL   Microbiology 148 (Pt 11), 3671-3680 (2002)
   PUBMED   12427957
REFERENCE   2  (bases 1 to 4931)
  AUTHORS   Chanto,G., Occhialini,A., Gras,N., Megraud,F. and Marais,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-DEC-2000) Laboratoire de Bacteriologie, Universite
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1..4931
                     /organism="Helicobacter pylori"
                     /mol_type="genomic DNA"
                     /strain="CR35"
                     /specific_host="Homo sapiens"
                     /db_xref="taxon:210"
                     /map="region F between JHP43/HP50 and JHP47/HP55"
                     /note="isolated from patient with chronic gastritis"
     CDS             complement(1..1086)
                     /note="HP0051-like protein"
                     /codon_start=1
                     /transl_table=11
                     /product="cytosine-specific DNA methyltransferase"
                     /protein_id="AAL37451.1"
                     /db_xref="GI:17225522"
                     /translation="MNYKILDLFCGAGGFSAGLGYLKEFDALIGLDCDKQALITFENN
                     HKNAIGVCGDITQTEIKEKVIELAKKLEINMIIGGPPCQGFSNKGKNLGLKDPRNFLF
                     LEYIEIVKAIKPEIFIIENVKNLISCAKGYFLEEIKERLNALGYQLSYQILNAKDYGV
                     PQNRERAFIVGASRFSFDFNLLEPSQSVNVQDAISDLAYLCSNEGAFESDYLNPIQSS
                     YQALMRKDSPKLYNHQATNHSQAALEKLKLINKEQGKECLPKNLHGKQQFKSTWGRLN
                     WNKISPTIDTRFDTPSNGTNSHPELHRSITPREAARIQSFSDDYIFYGNKTSVCKQIG
                     NAVPPLLALALGKAILKSARNDTNLSR"
     CDS             complement(1083..2309)
                     /codon_start=1
                     /transl_table=11
                     /product="putative HP0052-like protein"
                     /protein_id="AAL37452.1"
                     /db_xref="GI:17225523"
                     /translation="MLVSRFLNDDIDPFNLGVLLSRFQIKNGCIYGVCSYKASKFVPD
                     YEKSKARVLDALNTLSVHPIWQSNQESVTKIKGTFVFILEKDLHLDENAFYKKLLNLI
                     IDNDFFNRSHSMTPNQKRFLSGFFESRGSIDTQRNFLTSDYFFHSPLEFNKFHYLIDH
                     FNIPSEALNFNFRELQPEHAQGINQRNAQFRIYLNWYLHHIGLFNPYKARIAEHVFKT
                     TLIYDGIYHKLSYPPTTKYHGNSFTERAHFYLKNVYQQDLDDKSIEKLREQLGWIQKS
                     EEFKRDSKIINFYRISTPNVCSACCGDYDIKERSFISLPLYKITQNPNSYYTEIHHVV
                     SLGKNKELDVLENLAKLCPTCHRALKKGSSEEEFQKRLIRNILNHNKDNLEFAQLRFE
                     TDDFLTLIDKIHESLK"
     CDS             complement(2676..3707)
                     /function="methyltransferase"
                     /note="JHP45-like protein"
                     /codon_start=1
                     /transl_table=11
                     /product="type II DNA modification enzyme"
                     /protein_id="AAL37453.1"
                     /db_xref="GI:17225524"
                     /translation="MQNLLIQAENAIALLFLLNDKNLKGKIDLIYIDPPFATNNHFTI
                     TNGRATTISNSKNGDIAYSDKVVGMDFIEFLKQRLVLLKELLSEQGSIYVHTDCKIGH
                     YVKVMLDEIFGIQNFRNEITRIKCNPKNFKRIGYGNIKDMILFYSKGKNPIFNEPKIP
                     YTPQDLEKRFPKIDKDKRRYTTVPIHAPGEVESGECSKAFKGMLPPKGRHWRTDIATL
                     ERWDKEGLIEYSNNNNPRKKIYALEQAGKRVQDIWEFKDPQYPSYPTEKNAQLLDLII
                     KTSSNKGSIVLDCFCGSGTTLKSAFLLQRKFIGIDNSNLAIQACKNKLETITKDLFVS
                     QNFYDFLVF"
     CDS             complement(3717..4502)
                     /note="JHP46-like protein"
                     /codon_start=1
                     /transl_table=11
                     /product="type II restriction enzyme"
                     /protein_id="AAL37454.1"
                     /db_xref="GI:17225525"
                     /translation="MVDYWTKLSIEYANQRSYLDDLFQVYPTIPEGLREIDSKIWSNV
                     EYHFKRKDNLALITELLNLDLFPIKDSYMAYLKRDKSALERNPRTINRICGRLYEMGL
                     NKIFEKCSEPKETNRQIGPMFKDWINNKSLGVEPVDLNDFIANENDAILRASDNIMAE
                     FAKSHLNYHHHKGLDFVARFNKKYIIGEAKFLTDFGGHQNAQFNDAISTIEAPNIKAI
                     KVAILDGVLYIESNNKMRKLLDTTYRNYNIMSALVLREFLYQI"
ORIGIN

  Query Match              3.1%;  Score 179.2;  DB 1;  Length 4931;
  Best Local Similarity   51.5%;  Pred. No. 1.3e-27;
  Matches  471;  Conservative    0;  Mismatches  428;  Indels   15;  Gaps    2;

Qy       1664 ACCCTTCTAGTGAAGGCTTTGACCGCCTCCCAGGAGGCATCTATGCCGATGGATCGCCGC 1723
Db       2737 AGCTTGTTTTTGCAAGCTTGGATAGCTAAATTGGAATTATCAATGCCTATAAATTTTCGT 2796

Qy       1724 TTTAAGAGGGGTGAGGCTATAAGCGTAGTACCGGAGCCTGCGAAGGGATCGAGCACTAAA 1783
Db       2797 TGCAATAAAAACGCAGATTTTAAGGTTGTTCCAGAACCACAAAAACAATCTAAAACAATG 2856

Qy       1784 TCCCCCTCGTTACTCCCTGTTTGGACGATGAGCTTGAGCATGTCCAGATTTTTCTCGGTG 1843
Db       2857 CTACCTTTATTAGAAGAGGTTTTAATGATTAAGTCTAATAATTGAGCGTTTTTTTCTGTG 2916

Qy       1844 GGGTATCGCGGGTACGGAGGATCCTTGAACTGCCAAACGTCCTGGAGCTTCTTCCCCTTC 1903
Db       2917 GGGTAGCTTGGATATTGTGGGTCTTTAAATTCCCAGATGTCTTGGACTCTTTTGCCAGCT 2976

Qy       1904 TTCAGGCGATCCCGAGCGTAAACTTTCTTCCGCGGCACCCCGTTCTTTGACCAGACAATA 1963
Db       2977 T------GTTCTAAGGCATAAATTTTTTTTCTAGGGTTATTATTGTTAGAATACTCAATC 3030

Qy       1964 AGCCCTTGAGCGTCTAGCTCGTCAAGCTTCTCCGGGGGATAGCGCCAATGCCGTCCAGGA 2023
Db       3031 AAACCTTCTTTATCCCAACGCTCAAGTGTGGCAATATCAGTGCGCCAATGCCGCCCTTTT 3090

Qy       2024 GGGGGAAGTATTCCTCGCCAAGGCCTTCCGGTAGGGCCATCCTTGGTTTCTCCAGGAGCA 2083
Db       3091 GGAGGTAGCATACCTTTAAATGCTTTAGAGCATTCGCCACTTTCTACTTCTCCTGGAGCA 3150

Qy       2084 TGCAGGGGATTGGTGGTGTACCGTTCCCCGTTCTCGTCTACAAAGGGGAAAAGCCTAGCG 2143
Db       3151 TGTATTGGAACGGTAGTGTAACGCCTTTTATCCTTGTCAATTTTAGGGAATCGTTTTTCT 3210

Qy       2144 ATCTCCTCTTCCGAATAGGGGCTAGCCGATTCGTTCCAAACGTAGTCCCGCGTTTTGGAG 2203
Db       3211 AAATCTTGTGGCGTATAAGGAATCTTAGGCTCGTTAAAAATGGGATTTTTCCTTTAGAA 3270

Qy       2204 TAGACGAGGATCATGTCCTTTTGCGATCCGAAGGCCTTACGGGAAAAGTTTTTGGGATTT 2263
Db       3271 TAAAATAAAATCATATCTTTTATGTTACCATAGCCTATCCTTTTAAAATTTTTAGGATTG 3330

Qy       2264 GAAGCGATGCGGGCGATATGGTTAACGAAGTTTCGCCGGCCAAAGACCTCATCAAGGATG 2323
Db       3331 CACTTTATACGTGTGATTTCATTTCTAAAATTTTGTATGCCAAATATTTCATCTAACATC 3390

Qy       2324 AGCTTCACCTCGAACCCGTATTTCTCGTCTATGTGAACGAAGATCAGTCCTGAGTCCGCC 2383
Db       3391 ACCTTAACATAATGCCCTATCTTACAATCTGTATGCACATAGATAGAGCCTTGTTCTGAA 3450

Qy       2384 ATCAGCTCCCTGAGAAGTATCAAGCGCTCCCTCAGGAACTCCACAAACTGAGGACCATCG 2443
Db       3451 AGCAATTCTTTAAGCAATACCAAGCGTTGTTTTAAAAATTCTATAAAATCCATACCCACT 3510
```

```
Qy       2444 AGGGTGTCATCGTAGCCCAACTGACCGTTTTTGGGCTGGCTGACGGTAGCAACGCGATCT 2503
             |    ||||    ||  | |  |  || || || |  | ||| |  || |   | |   |
Db       3511 ACTTTGTCGCTATAAGCAATATCGCCATTCTTAGAATTGCTAATTGTGGTTGCTCTGCCA 3570

Qy       2504 GTTTCATCGCCGCCAACGAGAAACTGCTGGCCGGTTCCATAAGGCGGGTCAATATAGACC 2563
              ||           |    ||| || | |   ||  || | || ||||||||||| | |
Db       3571 TTT---------GTGATAGTAAAATGATTGTTTGTAGCAAATGGAGGGTCAATATATATC 3621

Qy       2564 AACTGGACCTTCCC 2577
              || |  |  || ||
Db       3622 AAATCTATTTTTCC 3635
```

Search completed: March 18, 2004, 02:39:47
Job time : 14877.6 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: March 17, 2004, 20:20:31 ; Search time 1429.28 Seconds
(without alignments)
17384.739 Million cell updates/sec

Title: US-09-664-186-6
Perfect score: 5849
Sequence: 1 tctagaaggtcagggtggac.....tgtagaggaccagaagaccc 5849

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	5849	100.0	5849	5	AAD04668	Aad04668 Thermus p
c	2	1242	21.2	1242	2	AAX01848	Aax01848 Thermus s
c	3	1026	17.5	1026	5	AAD04666	Aad04666 Thermus r
	4	999	17.1	999	2	AAX01849	Aax01849 Thermus s
c	5	185.6	3.2	1032	2	AAT67772	Aat67772 H. pylori
c	6	185.6	3.2	1032	2	AAT77452	Aat77452 H. pylori
c	7	180	3.1	180	5	AAD04667	Aad04667 Thermus p
c	8	59	1.0	2000	7	ADA71938	Ada71938 Rice gene
	9	57.6	1.0	2000	7	ADA71938	Ada71938 Rice gene
c	10	56.4	1.0	2030	7	ADA70342	Ada70342 Rice gene
c	11	56.4	1.0	2031	9	ADC08176	Adc08176 Rice DNA
c	12	55.4	0.9	2028	8	ADA48615	Ada48615 Rice gene
c	13	55.4	0.9	2028	9	ADC08212	Adc08212 Rice DNA
c	14	54.8	0.9	2028	7	ADA70341	Ada70341 Rice gene
c	15	54.8	0.9	2060	7	ADA71232	Ada71232 Rice gene
c	16	54.4	0.9	615	2	AAQ37036	Aaq37036 SOD-T gen
c	17	53.2	0.9	1755	7	ADA70624	Ada70624 Rice gene
c	18	53.2	0.9	1965	8	ADA48223	Ada48223 Rice gene
c	19	53.2	0.9	1965	9	ADC08174	Adc08174 Rice DNA
c	20	52.4	0.9	113193	7	AAD54645	Aad54645 Streptomy
c	21	51.6	0.9	1815	5	AAD09400	Aad09400 Zea mays
	22	50.2	0.9	509	8	ACL23248	Acl23248 DNA clone
c	23	49.6	0.8	1674	7	ACA38419	Aca38419 Prokaryot
c	24	49.6	0.8	1677	7	ACA40670	Aca40670 Prokaryot
	25	49.6	0.8	110000	4	AAI99682_27	Continuation (28 o
	26	49.6	0.8	110000	4	AAI99683_27	Continuation (28 o
	27	49.4	0.8	349980	6	ABQ81846	Abq81846 Bifidobac
c	28	48.8	0.8	1236	9	ADC36241	Adc36241 Weed cont
c	29	48.8	0.8	1454	9	ADC36242	Adc36242 Weed cont
	30	48.6	0.8	21034	2	AAV62154	Aav62154 HSV-2 str
	31	48.6	0.8	117213	2	AAV62176	Aav62176 HSV-2 str
	32	48.6	0.8	154746	6	AAD25519	Aad25519 Human her
c	33	47.8	0.8	963	7	ACA24030	Aca24030 Prokaryot
	34	47.4	0.8	10732	3	AAA10594	Aaa10594 Gene enco
c	35	47.2	0.8	1194	7	ADA69660	Ada69660 Rice gene
c	36	47	0.8	6094	3	AAA71428	Aaa71428 Expressio
c	37	46.8	0.8	1605	7	ABT31955	Abt31955 High-affi
c	38	46.6	0.8	924	6	ABK49877	Abk49877 Environme
c	39	46.6	0.8	7659	7	ACD13414	Acd13414 Human DNA
c	40	46.6	0.8	10065	7	AAL54465	Aal54465 Human cad
c	41	46.6	0.8	10065	8	ACA63037	Aca63037 Human Cad
c	42	46.6	0.8	10455	7	AAL54468	Aal54468 Human cad
c	43	46.6	0.8	10455	8	ACA63040	Aca63040 Human Cad
c	44	46.6	0.8	11073	7	AAL54466	Aal54466 Human cad
c	45	46.6	0.8	11073	8	ACA63038	Aca63038 Human Cad

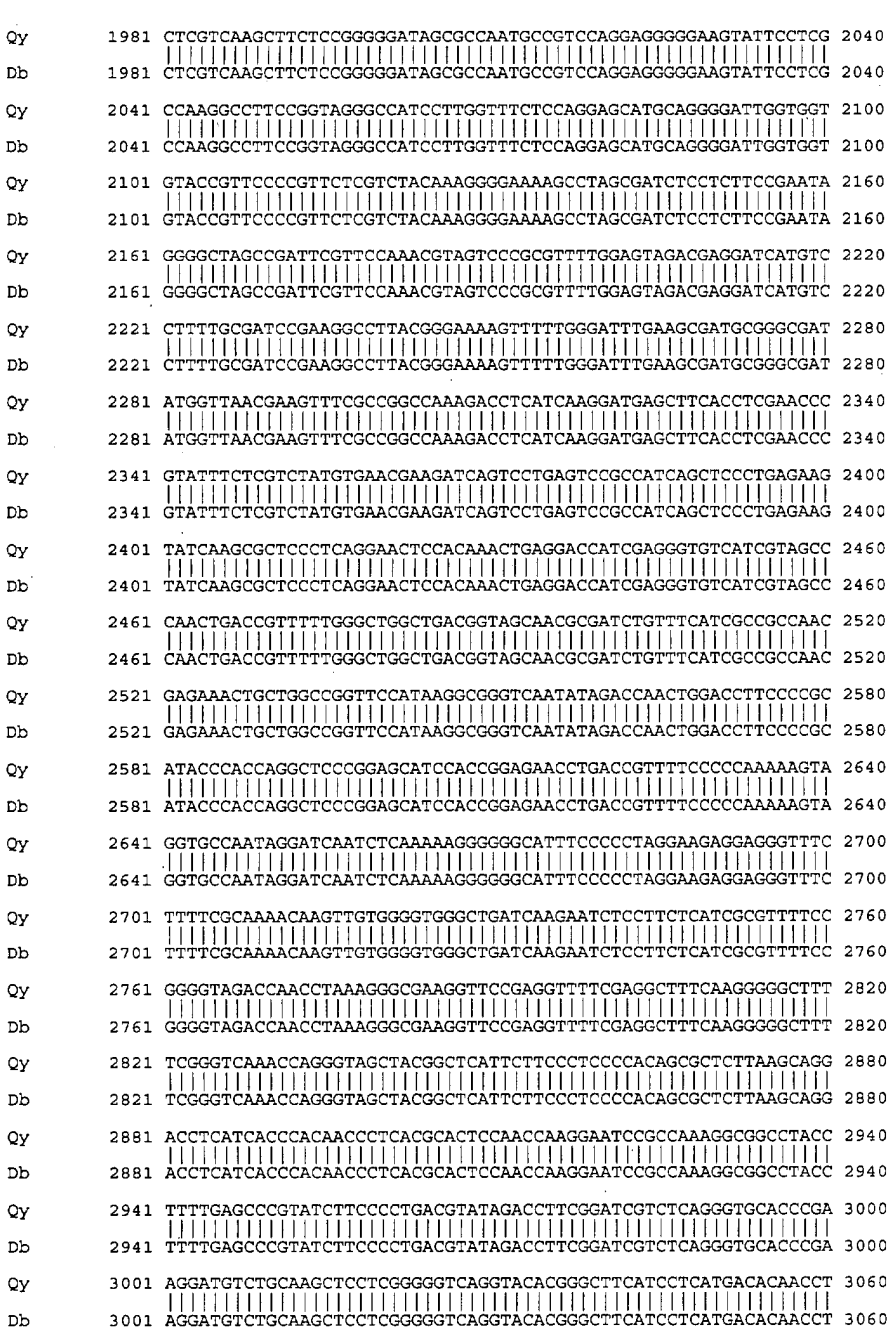
ALIGNMENTS

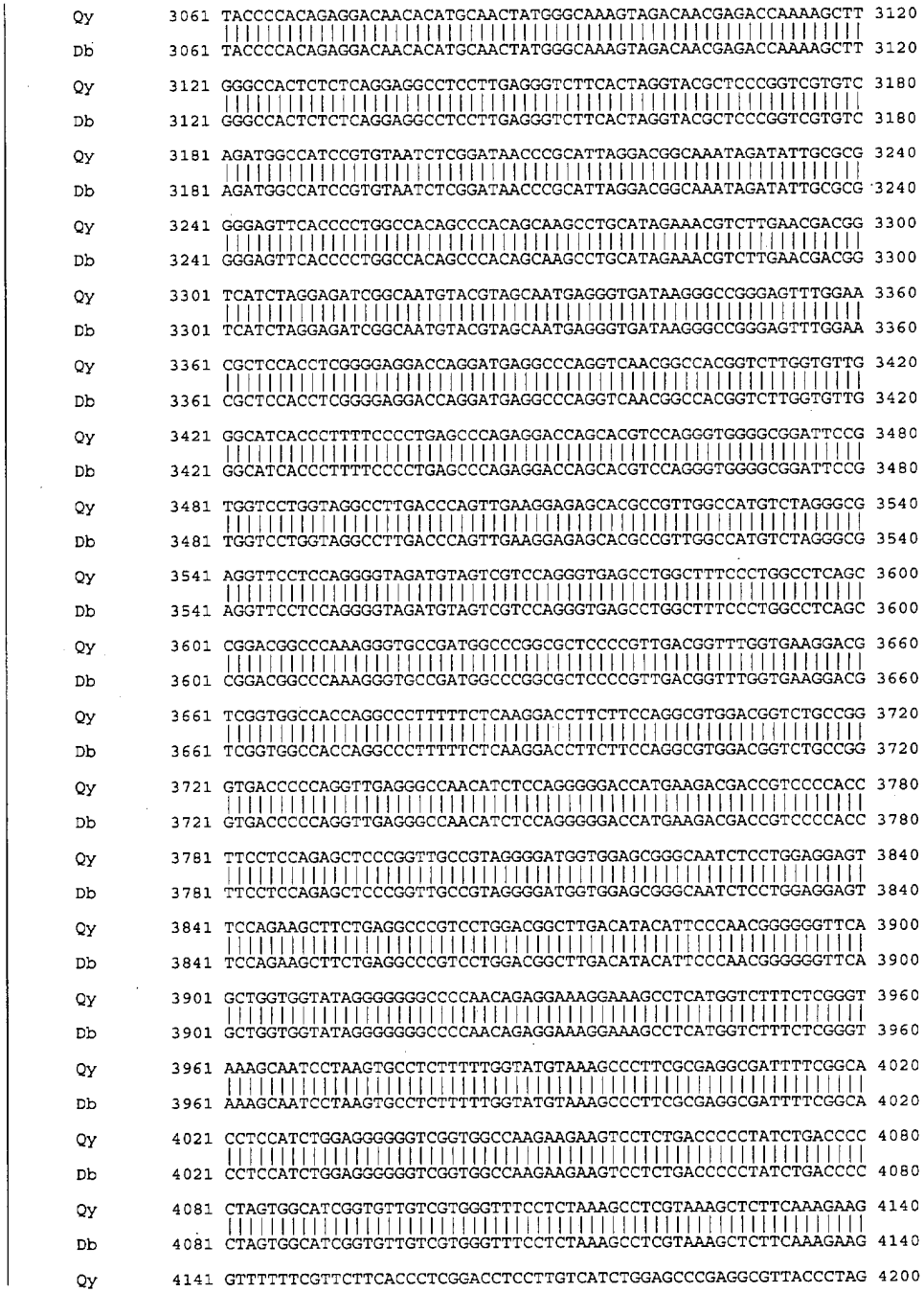
RESULT 1
AAD04668
ID AAD04668 standard; DNA; 5849 BP.
XX
AC AAD04668;
XX
DT 04-JUL-2001 (first entry)
XX
DE Thermus plasmid pTsp45S DNA sequence.
XX
KW Replication protein; RepT; partition protein; ParA; pTsp45S plasmid;
KW kanamycin-resistance gene; thermophilic transformation; Ori;
KW replication origin; ds.
XX
OS Thermus sp.
XX
PN US6207377-B1.
XX
PD 27-MAR-2001.
XX
PF 14-AUG-1998; 98US-00134246.
XX
PR 14-AUG-1998; 98US-00134246.
XX
PA (NEWE) NEW ENGLAND BIOLABS INC.
XX
PI Wayne J, Xu S;
XX
DR WPI; 2001-298939/31.
XX
PT Cloning Thermus species (Ts) plasmid genes comprises transforming
PT Escherichia coli with cloned recombinant plasmid containing Ts and E.coli
PT origins of replication, isolating cloned recombinant plasmid from E.coli
PT and transforming Ts cell.
XX
PS Example I; Fig 3; 32pp; English.
XX
CC The present sequence is Thermus plasmid pTsp45S DNA. The open reading
CC frame of pTsp45S plasmid is the replication protein, RepT which is needed
CC for thermophilic plasmid replication. The invention relates to Thermus
CC sp. replication protein RepT, partition protein ParA and their
CC corresponding DNA molecules which relates to recombinant DNA molecules
CC encoding plasmid DNA replication origins in Thermus, as well as to
CC shuttle vectors which contain the same. The invention also relates to
CC method useful for cloning Thermus sp. plasmid genes which comprises
CC inserting plasmid DNA comprising a Thermus sp. origin of replication

CC (Ori) into a recombinant plasmid comprising a thermostable kanamycin-
CC resistance gene and an Escherichia coli Ori, to produce a cloned
CC recombinant plasmid. This cloned recombinant plasmid is transformed with
CC an E. coli. host cell, and E. coli. host cell cultured for the expression
CC of cloned recombinant plasmid. The cloned recombinant plasmid isolated
CC from E. coli host cell is then transformed with Thermus sp. host cell and
CC Thermus sp. host cell is cultured. Thus Thermus sp. plasmid genes are
CC cloned. These plasmid DNAs are used for thermophilic transformation
XX
SQ Sequence 5849 BP; 1377 A; 1501 C; 1670 G; 1301 T; 0 U; 0 Other;

Query Match 100.0%; Score 5849; DB 5; Length 5849;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 5849; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 TCTAGAAGGTCAGGGTGGACAAGGAAAACACCATAGCCCCTGCCAAGAAGATGGACGAGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TCTAGAAGGTCAGGGTGGACAAGGAAAACACCATAGCCCCTGCCAAGAAGATGGACGAGT 60

Qy         61 TGGTGTCCGGAAAAGTGGCCATCCGGGGCGCTCTTGACAACTATTTTCCAGCGGTGGCCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGGTGTCCGGAAAAGTGGCCATCCGGGGCGCTCTTGACAACTATTTTCCAGCGGTGGCCA 120

Qy        121 CCGGCATTGGCCACGAGGTACGAGCTTGTGGAGTAGACGGCCACAAAGGGGTCGTCCTCA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGCATTGGCCACGAGGTACGAGCTTGTGGAGTAGACGGCCACAAAGGGGTCGTCCTCA 180

Qy        181 AACTTCTTTTCTAGTGCCGCTTGGACGAAGGGGAGGAAGAGGAAAGGCTTCATGGCCTCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AACTTCTTTTCTAGTGCCGCTTGGACGAAGGGGAGGAAGAGGAAAGGCTTCATGGCCTCA 240

Qy        241 CCTCCTTCCCCTCCTCCTTGGCGGCCTTAGCGGCGTAAAACTCTGAGACGGCCTGAAGTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCTCCTTCCCCTCCTCCTTGGCGGCCTTAGCGGCGTAAAACTCTGAGACGGCCTGAAGTT 300

Qy        301 TAGGGATTTCGCTTTCGGGGATAAGAATCCGGCGGCTCAGGGGATGCCGGATGGCCCTTA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TAGGGATTTCGCTTTCGGGGATAAGAATCCGGCGGCTCAGGGGATGCCGGATGGCCCTTA 360

Qy        361 TCCTGCCGTCCCTTATGTACTCGTAAATGGTGGCCTTGGGTACTTTAAACCGTTCTGAAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCCTGCCGTCCCTTATGTACTCGTAAATGGTGGCCTTGGGTACTTTAAACCGTTCTGAAA 420

Qy        421 CTTCTCTAACAGAGAGCACAAAACCTCTAAAAACCTATCAATCCCACCGATTCCAGTATA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTTCTCTAACAGAGAGCACAAAACCTCTAAAAACCTATCAATCCCACCGATTCCAGTATA 480

Qy        481 CCATAAATGGCACAAAGTTTTGAGAAGGTGGTCAAACAAAAAGGCTTTCTCGGTCAGGTT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCATAAATGGCACAAAGTTTTGAGAAGGTGGTCAAACAAAAAGGCTTTCTCGGTCAGGTT 540

Qy        541 ATGGTGAGGTGGGGGCGGTCAAAGGCCGACTTAAGTTTGGTAAAGCCGGGAGGAAGCAAA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATGGTGAGGTGGGGGCGGTCAAAGGCCGACTTAAGTTTGGTAAAGCCGGGAGGAAGCAAA 600

Qy        601 CCGGGGTGTTACCATGCAACAGATGGCCGAGTGGAACGTGTGGACACAGAGAAGCGTTGA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCGGGGTGTTACCATGCAACAGATGGCCGAGTGGAACGTGTGGACACAGAGAAGCGTTGA 660

Qy        661 GCTTCTGGAGAAGGGGTATTTGGATAAACTACTGCAGGTCTATAAAGGGGAAAGTGGCTC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GCTTCTGGAGAAGGGGTATTTGGATAAACTACTGCAGGTCTATAAAGGGGAAAGTGGCTC 720

Qy        721 TTCGAGGTCAGTACCAGAGGAGGTAGAGGAAAAACTTCGCGAGGCCTACAAGGCATACGA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TTCGAGGTCAGTACCAGAGGAGGTAGAGGAAAAACTTCGCGAGGCCTACAAGGCATACGA 780

Qy        781 GGGGAGGCAGGATAGTCCGGAGGCAGAAACGAAACTCGTGGAAGCCGTGCTAAATGCCAG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GGGGAGGCAGGATAGTCCGGAGGCAGAAACGAAACTCGTGGAAGCCGTGCTAAATGCCAG 840

Qy        841 AAAAAAGGTCGAGCGGTCCCCCTTCAATCACCCCTACCTGCCTTTGGTCTACTACCTGGT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAAAAAGGTCGAGCGGTCCCCCTTCAATCACCCCTACCTGCCTTTGGTCTACTACCTGGT 900

Qy        901 TTCGGAAAAAGCAGAAAAAGCGAACAAGGCCCTTGAGGAGGCATTGCAGGAGGTTGCCTC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCGGAAAAAGCAGAAAAAGCGAACAAGGCCCTTGAGGAGGCATTGCAGGAGGTTGCCTC 960

Qy        961 AAAGCACCCAGAAACCATCCGCGTCCTGGCCAAGGAAGCGCAAAGAAGAGGCGTAGAAGC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 AAAGCACCCAGAAACCATCCGCGTCCTGGCCAAGGAAGCGCAAAGAAGAGGCGTAGAAGC 1020

Qy       1021 CTTGATCCAAAGGCTCAAGGAGCCTCCCGAAATAAATCGGCAGATAGGGCCGATGTTCAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CTTGATCCAAAGGCTCAAGGAGCCTCCCGAAATAAATCGGCAGATAGGGCCGATGTTCAA 1080

Qy       1081 AAGGTGGTACAAAGAAGAGCTAAAGGGGAAAATAGAAGAGAGGCTTCCAGGCCCTACCAA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AAGGTGGTACAAAGAAGAGCTAAAGGGGAAAATAGAAGAGAGGCTTCCAGGCCCTACCAA 1140

Qy       1141 ACCAAAGATTGTGGTAGTATCCCCTGAAAAAAGTAAACCGGAGCAAGCACCCCTTATTGC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ACCAAAGATTGTGGTAGTATCCCCTGAAAAAAGTAAACCGGAGCAAGCACCCCTTATTGC 1200

Qy       1201 GGAGAGAGAAGCGGGCATCATCATATACACGGGATCGGATGAAGCTTTGAAAGATGCCGC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GGAGAGAGAAGCGGGCATCATCATATACACGGGATCGGATGAAGCTTTGAAAGATGCCGC 1260

Qy       1261 CAAGGAAAACCTGGGCCTTGGCGAGGAAGCAGAACTAGGCACCAAGGGCGTAGATTTCTA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CAAGGAAAACCTGGGCCTTGGCGAGGAAGCAGAACTAGGCACCAAGGGCGTAGATTTCTA 1320

Qy       1321 CGTGGTCATCCGGCGTAGCCCTGAAGAGACATGGCACCTAACAGGAGAAGTGAAGTTTCA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CGTGGTCATCCGGCGTAGCCCTGAAGAGACATGGCACCTAACAGGAGAAGTGAAGTTTCA 1380

Qy       1381 ATCCGACTTTGGCGGAAACCAAGACAACCAGAAACTAGTAGCAAAGGCTTCCATAAGGTT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 ATCCGACTTTGGCGGAAACCAAGACAACCAGAAACTAGTAGCAAAGGCTTCCATAAGGTT 1440

Qy       1441 GGACCTTGAGAAGAGGCACATAGGAATAGTGGTGGTGGACGGAATGCCTGTGGTGAGCAA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 GGACCTTGAGAAGAGGCACATAGGAATAGTGGTGGTGGACGGAATGCCTGTGGTGAGCAA 1500

Qy       1501 GTTTCGTGGGTGGGCCGGACTGGGGAAAGAAACGATCGTTACATCCGTACTCCTCCTTCC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GTTTCGTGGGTGGGCCGGACTGGGGAAAGAAACGATCGTTACATCCGTACTCCTCCTTCC 1560

Qy       1561 AGACCTGATAGCGGAGCTCTACCAAAAGGGTGAAGAAGCCCTGGGCCTCTAGAAGGCGGA 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 AGACCTGATAGCGGAGCTCTACCAAAAGGGTGAAGAAGCCCTGGGCCTCTAGAAGGCGGA 1620

Qy       1621 CACAATCTCAAACTTGTGCTGTAGCCTGGGGAAATCCTCTAACACCCTTCTAGTGAAGGC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 CACAATCTCAAACTTGTGCTGTAGCCTGGGGAAATCCTCTAACACCCTTCTAGTGAAGGC 1680

Qy       1681 TTTGACCGCCTCCCAGGAGGCATCTATGCCGATGGATCGCCGCTTTAAGAGGGGTGAGGC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TTTGACCGCCTCCCAGGAGGCATCTATGCCGATGGATCGCCGCTTTAAGAGGGGTGAGGC 1740

Qy       1741 TATAAGCGTAGTACCGGAGCCTGCGAAGGGATCGAGCACTAAATCCCCCTCGTTACTCCC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 TATAAGCGTAGTACCGGAGCCTGCGAAGGGATCGAGCACTAAATCCCCCTCGTTACTCCC 1800

Qy       1801 TGTTTGGACGATGAGCTTGAGCATGTCCAGATTTTTCTCGGTGGGGTATCGCGGGTACGG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 TGTTTGGACGATGAGCTTGAGCATGTCCAGATTTTTCTCGGTGGGGTATCGCGGGTACGG 1860

Qy       1861 AGGATCCTTGAACTGCCAAACGTCCTGGAGCTTCTTCCCCTTCTTCAGGCGATCCCGAGC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 AGGATCCTTGAACTGCCAAACGTCCTGGAGCTTCTTCCCCTTCTTCAGGCGATCCCGAGC 1920

Qy       1921 GTAAACTTTCTTCCGCGGCACCCCGTTCTTTGACCAGACAATAAGCCCTTGAGCGTCTAG 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 GTAAACTTTCTTCCGCGGCACCCCGTTCTTTGACCAGACAATAAGCCCTTGAGCGTCTAG 1980
```


Qy 1981 CTCGTCAAGCTTCTCCGGGGGATAGCGCCAATGCCGTCCAGGAGGGGGAAGTATTCCTCG 2040
Db 1981 CTCGTCAAGCTTCTCCGGGGGATAGCGCCAATGCCGTCCAGGAGGGGGAAGTATTCCTCG 2040
Qy 2041 CCAAGGCCTTCCGGTAGGGCCATCCTTGGTTTCTCCAGGAGCATGCAGGGGATTGGTGGT 2100
Db 2041 CCAAGGCCTTCCGGTAGGGCCATCCTTGGTTTCTCCAGGAGCATGCAGGGGATTGGTGGT 2100
Qy 2101 GTACCGTTCCCCGTTCTCGTCTACAAAGGGGAAAAGCCTAGCGATCTCCTCTTCCGAATA 2160
Db 2101 GTACCGTTCCCCGTTCTCGTCTACAAAGGGGAAAAGCCTAGCGATCTCCTCTTCCGAATA 2160
Qy 2161 GGGGCTAGCCGATTCGTTCCAAACGTAGTCCCGCGTTTTGGAGTAGACGAGGATCATGTC 2220
Db 2161 GGGGCTAGCCGATTCGTTCCAAACGTAGTCCCGCGTTTTGGAGTAGACGAGGATCATGTC 2220
Qy 2221 CTTTTGCGATCCGAAGGCCTTACGGGAAAAGTTTTTGGGATTTGAAGCGATGCGGGCGAT 2280
Db 2221 CTTTTGCGATCCGAAGGCCTTACGGGAAAAGTTTTTGGGATTTGAAGCGATGCGGGCGAT 2280
Qy 2281 ATGGTTAACGAAGTTTCGCCGGCCAAAGACCTCATCAAGGATGAGCTTCACCTCGAACCC 2340
Db 2281 ATGGTTAACGAAGTTTCGCCGGCCAAAGACCTCATCAAGGATGAGCTTCACCTCGAACCC 2340
Qy 2341 GTATTTCTCGTCTATGTGAACGAAGATCAGTCCTGAGTCCGCCATCAGCTCCCTGAGAAG 2400
Db 2341 GTATTTCTCGTCTATGTGAACGAAGATCAGTCCTGAGTCCGCCATCAGCTCCCTGAGAAG 2400
Qy 2401 TATCAAGCGCTCCCTCAGGAACTCCACAAACTGAGGACCATCGAGGGTGTCATCGTAGCC 2460
Db 2401 TATCAAGCGCTCCCTCAGGAACTCCACAAACTGAGGACCATCGAGGGTGTCATCGTAGCC 2460
Qy 2461 CAACTGACCGTTTTTGGGCTGGCTGACGGTAGCAACGCGATCTGTTTCATCGCCGCCAAC 2520
Db 2461 CAACTGACCGTTTTTGGGCTGGCTGACGGTAGCAACGCGATCTGTTTCATCGCCGCCAAC 2520
Qy 2521 GAGAAACTGCTGGCCGGTTCCATAAGGCGGGTCAATATAGACCAACTGGACCTTCCCCGC 2580
Db 2521 GAGAAACTGCTGGCCGGTTCCATAAGGCGGGTCAATATAGACCAACTGGACCTTCCCCGC 2580
Qy 2581 ATACCCACCAGGCTCCCGGAGCATCCACCGGAGAACCTGACCGTTTTCCCCCAAAAAGTA 2640
Db 2581 ATACCCACCAGGCTCCCGGAGCATCCACCGGAGAACCTGACCGTTTTCCCCCAAAAAGTA 2640
Qy 2641 GGTGCCAATAGGATCAATCTCAAAAAGGGGGGCATTTCCCCCTAGGAAGAGGAGGGTTTC 2700
Db 2641 GGTGCCAATAGGATCAATCTCAAAAAGGGGGGCATTTCCCCCTAGGAAGAGGAGGGTTTC 2700
Qy 2701 TTTTCGCAAAACAAGTTGTGGGGTGGGCTGATCAAGAATCTCCTTCTCATCGCGTTTTCC 2760
Db 2701 TTTTCGCAAAACAAGTTGTGGGGTGGGCTGATCAAGAATCTCCTTCTCATCGCGTTTTCC 2760
Qy 2761 GGGGTAGACCAACCTAAAGGGCGAAGGTTCCGAGGTTTTCGAGGCTTTCAAGGGGGCTTT 2820
Db 2761 GGGGTAGACCAACCTAAAGGGCGAAGGTTCCGAGGTTTTCGAGGCTTTCAAGGGGGCTTT 2820
Qy 2821 TCGGGTCAAACCAGGGTAGCTACGGCTCATTCTTCCCTCCCCACAGCGCTCTTAAGCAGG 2880
Db 2821 TCGGGTCAAACCAGGGTAGCTACGGCTCATTCTTCCCTCCCCACAGCGCTCTTAAGCAGG 2880
Qy 2881 ACCTCATCACCCACAACCCTCACGCACTCCAACCAAGGAATCCGCCAAAGGCGGCCTACC 2940
Db 2881 ACCTCATCACCCACAACCCTCACGCACTCCAACCAAGGAATCCGCCAAAGGCGGCCTACC 2940
Qy 2941 TTTTGAGCCCGTATCTTCCCCTGACGTATAGACCTTCGGATCGTCTCAGGGTGCACCCGA 3000
Db 2941 TTTTGAGCCCGTATCTTCCCCTGACGTATAGACCTTCGGATCGTCTCAGGGTGCACCCGA 3000
Qy 3001 AGGATGTCTGCAAGCTCCTCGGGGGTCAGGTACACGGGCTTCATCCTCATGACACAACCT 3060
Db 3001 AGGATGTCTGCAAGCTCCTCGGGGGTCAGGTACACGGGCTTCATCCTCATGACACAACCT 3060


Qy 3061 TACCCCACAGAGGACAACACATGCAACTATGGGCAAAGTAGACAACGAGACCAAAAGCTT 3120
Db 3061 TACCCCACAGAGGACAACACATGCAACTATGGGCAAAGTAGACAACGAGACCAAAAGCTT 3120
Qy 3121 GGGCCACTCTCTCAGGAGGCCTCCTTGAGGGTCTTCACTAGGTACGCTCCCGGTCGTGTC 3180
Db 3121 GGGCCACTCTCTCAGGAGGCCTCCTTGAGGGTCTTCACTAGGTACGCTCCCGGTCGTGTC 3180
Qy 3181 AGATGGCCATCCGTGTAATCTCGGATAACCCGCATTAGGACGGCAAATAGATATTGCGCG 3240
Db 3181 AGATGGCCATCCGTGTAATCTCGGATAACCCGCATTAGGACGGCAAATAGATATTGCGCG 3240
Qy 3241 GGGAGTTCACCCCTGGCCACAGCCCACAGCAAGCCTGCATAGAAACGTCTTGAACGACGG 3300
Db 3241 GGGAGTTCACCCCTGGCCACAGCCCACAGCAAGCCTGCATAGAAACGTCTTGAACGACGG 3300
Qy 3301 TCATCTAGGAGATCGGCAATGTACGTAGCAATGAGGGTGATAAGGGCCGGGAGTTTGGAA 3360
Db 3301 TCATCTAGGAGATCGGCAATGTACGTAGCAATGAGGGTGATAAGGGCCGGGAGTTTGGAA 3360
Qy 3361 CGCTCCACCTCGGGGAGGACCAGGATGAGGCCCAGGTCAACGGCCACGGTCTTGGTGTTG 3420
Db 3361 CGCTCCACCTCGGGGAGGACCAGGATGAGGCCCAGGTCAACGGCCACGGTCTTGGTGTTG 3420
Qy 3421 GGCATCACCCTTTTCCCCTGAGCCCAGAGGACCAGCACGTCCAGGGTGGGGCGGATTCCG 3480
Db 3421 GGCATCACCCTTTTCCCCTGAGCCCAGAGGACCAGCACGTCCAGGGTGGGGCGGATTCCG 3480
Qy 3481 TGGTCCTGGTAGGCCTTGACCCAGTTGAAGGAGAGCACGCCGTTGGCCATGTCTAGGGCG 3540
Db 3481 TGGTCCTGGTAGGCCTTGACCCAGTTGAAGGAGAGCACGCCGTTGGCCATGTCTAGGGCG 3540
Qy 3541 AGGTTCCTCCAGGGGTAGATGTAGTCGTCCAGGGTGAGCCTGGCTTTCCCTGGCCTCAGC 3600
Db 3541 AGGTTCCTCCAGGGGTAGATGTAGTCGTCCAGGGTGAGCCTGGCTTTCCCTGGCCTCAGC 3600
Qy 3601 CGGACGGCCCAAAGGGTGCCGATGGCCCGGCGCTCCCCGTTGACGGTTTGGTGAAGGACG 3660
Db 3601 CGGACGGCCCAAAGGGTGCCGATGGCCCGGCGCTCCCCGTTGACGGTTTGGTGAAGGACG 3660
Qy 3661 TCGGTGGCCACCAGGCCCTTTTTCTCAAGGACCTTCTTCCAGGCGTGGACGGTCTGCCGG 3720
Db 3661 TCGGTGGCCACCAGGCCCTTTTTCTCAAGGACCTTCTTCCAGGCGTGGACGGTCTGCCGG 3720
Qy 3721 GTGACCCCCAGGTTGAGGGCCAACATCTCCAGGGGGACCATGAAGACGACCGTCCCCACC 3780
Db 3721 GTGACCCCCAGGTTGAGGGCCAACATCTCCAGGGGGACCATGAAGACGACCGTCCCCACC 3780
Qy 3781 TTCCTCCAGAGCTCCCGGTTGCCGTAGGGGATGGTGGAGCGGGCAATCTCCTGGAGGAGT 3840
Db 3781 TTCCTCCAGAGCTCCCGGTTGCCGTAGGGGATGGTGGAGCGGGCAATCTCCTGGAGGAGT 3840
Qy 3841 TCCAGAAGCTTCTGAGGCCCGTCCTGGACGGCTTGACATACATTCCCAACGGGGGGTTCA 3900
Db 3841 TCCAGAAGCTTCTGAGGCCCGTCCTGGACGGCTTGACATACATTCCCAACGGGGGGTTCA 3900
Qy 3901 GCTGGTGGTATAGGGGGGGCCCCAACAGAGGAAAGGAAAGCCTCATGGTCTTTCTCGGGT 3960
Db 3901 GCTGGTGGTATAGGGGGGGCCCCAACAGAGGAAAGGAAAGCCTCATGGTCTTTCTCGGGT 3960
Qy 3961 AAAGCAATCCTAAGTGCCTCTTTTTGGTATGTAAAGCCCTTCGCGAGGCGATTTTCGGCA 4020
Db 3961 AAAGCAATCCTAAGTGCCTCTTTTTGGTATGTAAAGCCCTTCGCGAGGCGATTTTCGGCA 4020
Qy 4021 CCTCCATCTGGAGGGGGGTCGGTGGCCAAGAAGAAGTCCTCTGACCCCCTATCTGACCCC 4080
Db 4021 CCTCCATCTGGAGGGGGGTCGGTGGCCAAGAAGAAGTCCTCTGACCCCCTATCTGACCCC 4080
Qy 4081 CTAGTGGCATCGGTGTTGTCGTGGGTTTCCTCTAAAGCCTCGTAAAGCTCTTCAAAGAAG 4140
Db 4081 CTAGTGGCATCGGTGTTGTCGTGGGTTTCCTCTAAAGCCTCGTAAAGCTCTTCAAAGAAG 4140
Qy 4141 GTTTTTTCGTTCTTCACCCTCGGACCTCCTTGTCATCTGGAGCCCGAGGCGTTACCCTAG 4200

```
Db       4141 GTTTTTTCGTTCTTCACCCTCGGACCTCCTTGTCATCTGGAGCCCGAGGCGTTACCCTAG 4200
Qy       4201 GTCTTGGGGGTGATCCGGGGCAACCGCCTCGGTTTCGCCTTTTTATGGGTCCAAAATAAC 4260
Db       4201 GTCTTGGGGGTGATCCGGGGCAACCGCCTCGGTTTCGCCTTTTTATGGGTCCAAAATAAC 4260
Qy       4261 CGTCAGCCCAGCGGCTGGCAATCCCCCCTCCTAAAAGGCCGTTATAGGCCCTGCTAGGAG 4320
Db       4261 CGTCAGCCCAGCGGCTGGCAATCCCCCCTCCTAAAAGGCCGTTATAGGCCCTGCTAGGAG 4320
Qy       4321 GGGGGTAGTACTTTCCTACCCCCCTAGGCTTGGAGAGGCCTTAGGAGGTCTCCTAGGGCC 4380
Db       4321 GGGGGTAGTACTTTCCTACCCCCCTAGGCTTGGAGAGGCCTTAGGAGGTCTCCTAGGGCC 4380
Qy       4381 TCGTGGGGGTGTAGGGGTAACCTCATGGCCAGGCCGGCCGGCTCGGGACTCTGGAGGAGG 4440
Db       4381 TCGTGGGGGTGTAGGGGTAACCTCATGGCCAGGCCGGCCGGCTCGGGACTCTGGAGGAGG 4440
Qy       4441 CCTCCATAGCCTACTCGTGGTGGAGGTTTGTGAAGGGGTTCACTAATGCATACGGCTAGC 4500
Db       4441 CCTCCATAGCCTACTCGTGGTGGAGGTTTGTGAAGGGGTTCACTAATGCATACGGCTAGC 4500
Qy       4501 CTCGGGATCACGGCCAAATGGTATGCAGGTTTTGGTATAAAACCCTCAGGTTTGAGGCTA 4560
Db       4501 CTCGGGATCACGGCCAAATGGTATGCAGGTTTTGGTATAAAACCCTCAGGTTTGAGGCTA 4560
Qy       4561 GTTTATGTCGGTTTTATGCACCTTTGACTCGGATCACGGGCATAAACACCAGTTTCCTGC 4620
Db       4561 GTTTATGTCGGTTTTATGCACCTTTGACTCGGATCACGGGCATAAACACCAGTTTCCTGC 4620
Qy       4621 ACGAAAGAAAACTTTCGCGATCTAAGAGGGGGAAAGAGGTGTAGAGGGACGGCCTTCATG 4680
Db       4621 ACGAAAGAAAACTTTCGCGATCTAAGAGGGGGAAAGAGGTGTAGAGGGACGGCCTTCATG 4680
Qy       4681 AAAGTTGGCCTCTTAGGAGGCCGTTGTAGAGGGCCGTCTCGGGTTCAAATCCTTTCCCTC 4740
Db       4681 AAAGTTGGCCTCTTAGGAGGCCGTTGTAGAGGGCCGTCTCGGGTTCAAATCCTTTCCCTC 4740
Qy       4741 TCTCTCCAGGTTTCCGAGGTTCGAGGTCTTGGTCCAGGTCTTGTACCAAGTTTTTGACCA 4800
Db       4741 TCTCTCCAGGTTTCCGAGGTTCGAGGTCTTGGTCCAGGTCTTGTACCAAGTTTTTGACCA 4800
Qy       4801 AAGTCTATTCTCGGAATATAGGGGTATCTTGTCTATCTTCCCTACGGGATATCTCTGTCT 4860
Db       4801 AAGTCTATTCTCGGAATATAGGGGTATCTTGTCTATCTTCCCTACGGGATATCTCTGTCT 4860
Qy       4861 GTGTGAACTTGATCCCATCCCAATACATATCTCAATCTCCTAATCTCCTCTTCTCTCCAG 4920
Db       4861 GTGTGAACTTGATCCCATCCCAATACATATCTCAATCTCCTAATCTCCTCTTCTCTCCAG 4920
Qy       4921 ATCCCTAATCTCTTCTTCTACCTCTTTCTCCTCCCAATTAAGAATGGAGAGGAAAAACCC 4980
Db       4921 ATCCCTAATCTCTTCTTCTACCTCTTTCTCCTCCCAATTAAGAATGGAGAGGAAAAACCC 4980
Qy       4981 CGACCAGAACGAGCTTCTCGGGGTCAGTTTCGGTAATCTCGGGACAGGTTTTCATCGTCT 5040
Db       4981 CGACCAGAACGAGCTTCTCGGGGTCAGTTTCGGTAATCTCGGGACAGGTTTTCATCGTCT 5040
Qy       5041 AGGACGAGGATTAGGGCATGAAAAATGGGCTTTGACAAAATCTTTCTAAAAAATACTCCC 5100
Db       5041 AGGACGAGGATTAGGGCATGAAAAATGGGCTTTGACAAAATCTTTCTAAAAAATACTCCC 5100
Qy       5101 CGAGGTTGGGGAAGTGCCCTCGGGGAGAAGATTTTTGGCAGTTTAGATGTTATGCTCTAT 5160
Db       5101 CGAGGTTGGGGAAGTGCCCTCGGGGAGAAGATTTTTGGCAGTTTAGATGTTATGCTCTAT 5160
Qy       5161 CACGGGCCGGAGGCCTCCACGATAAGTTGTCTTGGCCAAGTACCGGGCCAGGTCGGGGGT 5220
Db       5161 CACGGGCCGGAGGCCTCCACGATAAGTTGTCTTGGCCAAGTACCGGGCCAGGTCGGGGGT 5220
Qy       5221 GCTCTTCAGCGTGGTGATGGTACTTTCACGGAAGTTCACAAGTCCTTTTAGAGGCTTCAG 5280
Db       5221 GCTCTTCAGCGTGGTGATGGTACTTTCACGGAAGTTCACAAGTCCTTTTAGAGGCTTCAG 5280
Qy       5281 GTCGGGGATAGTGCTCAAGTACTCCCAAGCGTTCTCGGGCCCGTGGTCGGGGAGAAGGAC 5340
Db       5281 GTCGGGGATAGTGCTCAAGTACTCCCAAGCGTTCTCGGGCCCGTGGTCGGGGAGAAGGAC 5340
Qy       5341 AAAGGGGTCGGGCAAAAGTTCATCTTTGTACTTAGGACGGATTACTTTAGCACCTGATAA 5400
Db       5341 AAAGGGGTCGGGCAAAAGTTCATCTTTGTACTTAGGACGGATTACTTTAGCACCTGATAA 5400
Qy       5401 CTTCAGGGCCGTTAAGAAGGGCCTCACCTCGGAGACGGGTGGAAGGAGGACGTGGGCGTG 5460
Db       5401 CTTCAGGGCCGTTAAGAAGGGCCTCACCTCGGAGACGGGTGGAAGGAGGACGTGGGCGTG 5460
Qy       5461 GAAGAAGACGAACCCCGATTTTTGGGAAGTCTCCCTCCAGTTTGATGATGAACGTTGGGA 5520
Db       5461 GAAGAAGACGAACCCCGATTTTTGGGAAGTCTCCCTCCAGTTTGATGATGAACGTTGGGA 5520
Qy       5521 GGAAGCCGGCCAGGATGTCTTTCATCGCGCCTCGAACCTCGGACACATAAAAAACTTTCG 5580
Db       5521 GGAAGCCGGCCAGGATGTCTTTCATCGCGCCTCGAACCTCGGACACATAAAAAACTTTCG 5580
Qy       5581 TGTTTGTCAGGGCAAGAGTGCTATGTATGAGGTAACCTTCGGGAGTACAAAGTGCCTCAA 5640
Db       5581 TGTTTGTCAGGGCAAGAGTGCTATGTATGAGGTAACCTTCGGGAGTACAAAGTGCCTCAA 5640
Qy       5641 GCCGCCTTTCCCAACGCTCCAAAACTCTAGGGTCAGGTGGTTTAGGTTTTCTGAAAAACT 5700
Db       5641 GCCGCCTTTCCCAACGCTCCAAAACTCTAGGGTCAGGTGGTTTAGGTTTTCTGAAAAACT 5700
Qy       5701 CTAGCTTTTCAGTGGTCATTCCTCACCCCTCTAGCACGTACTCTGGAAGGTAAACCTTTG 5760
Db       5701 CTAGCTTTTCAGTGGTCATTCCTCACCCCTCTAGCACGTACTCTGGAAGGTAAACCTTTG 5760
Qy       5761 ACACAGCGGCCAAGTCTAGCGTCTCCCAGTCCAGTTGGTCTGGGACGCGTGAGAAGGGGA 5820
Db       5761 ACACAGCGGCCAAGTCTAGCGTCTCCCAGTCCAGTTGGTCTGGGACGCGTGAGAAGGGGA 5820
Qy       5821 GGGGCTTGGTGTAGAGGACCAGAAGACCC 5849
Db       5821 GGGGCTTGGTGTAGAGGACCAGAAGACCC 5849
```

RESULT 2
AAX01848/c
ID AAX01848 standard; DNA; 1242 BP.
XX
AC AAX01848;
XX
DT 13-APR-1999 (first entry)
XX
DE Thermus sp. Tsp45IM genomic DNA.
XX
KW Thermostable enzyme; Tsp45IM; restriction endonuclease; methylase;
KW cleavage; ss.
XX
OS Thermus sp.
XX
PN US5866422-A.
XX
PD 02-FEB-1999.
XX
PF 29-OCT-1997; 97US-00960756.
XX
PR 29-OCT-1997; 97US-00960756.
XX
PA (NEWE) NEW ENGLAND BIOLABS INC.
XX
PI Xu S, Wayne J;
XX
DR WPI; 1999-141942/12.
DR P-PSDB; AAW92372.
XX

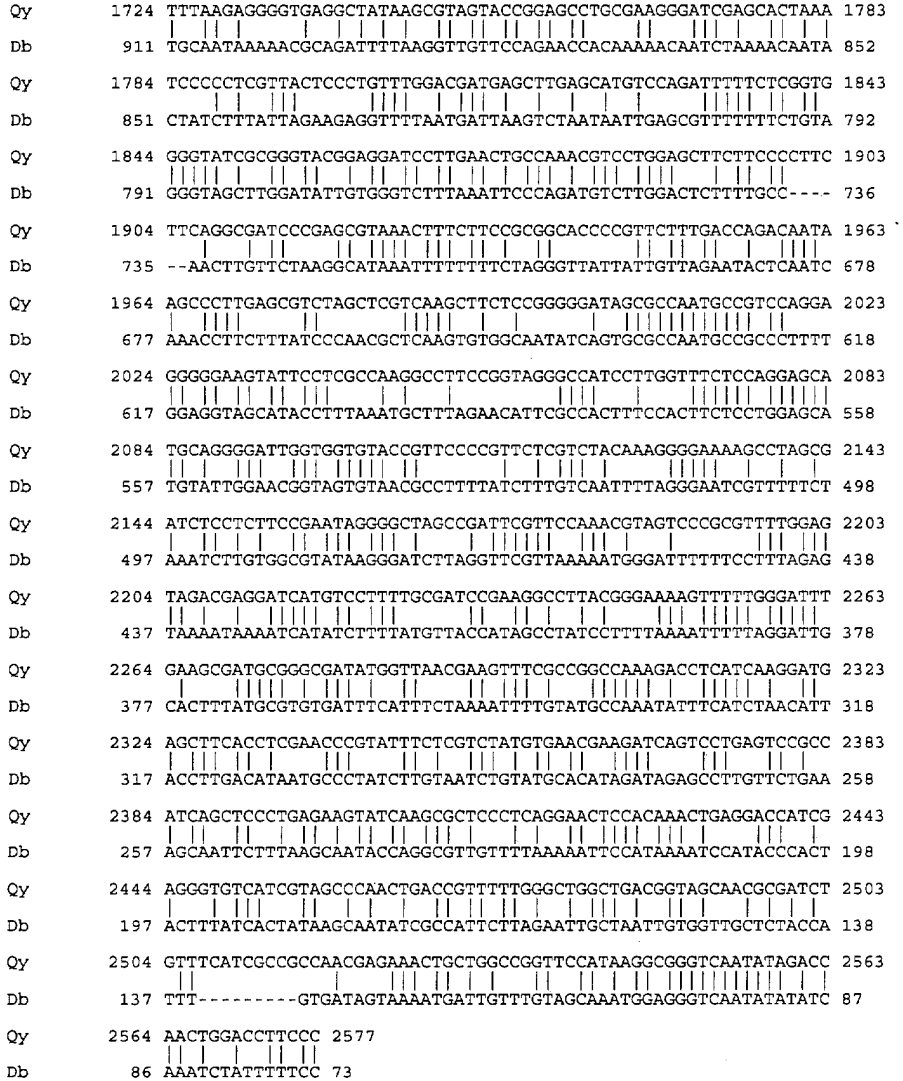
```
PT   DNA coding for Tsp45I restriction endonuclease and methylase from Thermus
PT   species YS45 - useful for recombinant production of thermostable enzyme
PT   in Escherichia coli.
XX
PS   Disclosure; Fig 1A-B; 22pp; English.
XX
CC   This sequence represents a novel restriction endonuclease methylase gene,
CC   Tsp45IM which has been isolated from Thermus sp. YS45. This DNA is useful
CC   for producing a thermostable endonuclease which can be used in a variety
CC   of assays, such as PCR, where high temperatures cannot be avoided. Tsp45I
CC   cleaves a double-stranded DNA molecule before the first G residue of the
CC   sequences 5'-GTGAC-3' and 5'-GTCAC-3'. The endonuclease creates sticky
CC   ends in the form of 5 nucleotide, single-stranded 5' overhangs. The
CC   production of recombinant Tsp45I in E. coli is simpler and produces 3
CC   multiply 105 units per gram of wet cell, about a ten-fold increase over
CC   that prepared from YS45
XX
SQ   Sequence 1242 BP; 291 A; 338 C; 336 G; 277 T; 0 U; 0 Other;

  Query Match             21.2%;  Score 1242;  DB 2;  Length 1242;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1242;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1609 CTAGAAGGCGGACACAATCTCAAACTTGTGCTGTAGCCTGGGGAAATCCTCTAACACCCT 1668
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1242 CTAGAAGGCGGACACAATCTCAAACTTGTGCTGTAGCCTGGGGAAATCCTCTAACACCCT 1183

Qy       1669 TCTAGTGAAGGCTTTGACCGCCTCCCAGGAGGCATCTATGCCGATGGATCGCCGCTTTAA 1728
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1182 TCTAGTGAAGGCTTTGACCGCCTCCCAGGAGGCATCTATGCCGATGGATCGCCGCTTTAA 1123

Qy       1729 GAGGGGTGAGGCTATAAGCGTAGTACCGGAGCCTGCGAAGGGATCGAGCACTAAATCCCC 1788
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1122 GAGGGGTGAGGCTATAAGCGTAGTACCGGAGCCTGCGAAGGGATCGAGCACTAAATCCCC 1063

Qy       1789 CTCGTTACTCCCTGTTTGGACGATGAGCTTGAGCATGTCCAGATTTTTCTCGGTGGGGTA 1848
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1062 CTCGTTACTCCCTGTTTGGACGATGAGCTTGAGCATGTCCAGATTTTTCTCGGTGGGGTA 1003

Qy       1849 TCGCGGGTACGGAGGATCCTTGAACTGCCAAACGTCCTGGAGCTTCTTCCCCTTCTTCAG 1908
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1002 TCGCGGGTACGGAGGATCCTTGAACTGCCAAACGTCCTGGAGCTTCTTCCCCTTCTTCAG 943

Qy       1909 GCGATCCCGAGCGTAAACTTTCTTCCGCGGCACCCCGTTCTTTGACCAGACAATAAGCCC 1968
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        942 GCGATCCCGAGCGTAAACTTTCTTCCGCGGCACCCCGTTCTTTGACCAGACAATAAGCCC 883

Qy       1969 TTGAGCGTCTAGCTCGTCAAGCTTCTCCGGGGGATAGCGCCAATGCCGTCCAGGAGGGGG 2028
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        882 TTGAGCGTCTAGCTCGTCAAGCTTCTCCGGGGGATAGCGCCAATGCCGTCCAGGAGGGGG 823

Qy       2029 AAGTATTCCTCGCCAAGGCCTTCCGGTAGGGCCATCCTTGGTTTCTCCAGGAGCATGCAG 2088
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        822 AAGTATTCCTCGCCAAGGCCTTCCGGTAGGGCCATCCTTGGTTTCTCCAGGAGCATGCAG 763

Qy       2089 GGGATTGGTGGTGTACCGTTCCCCGTTCTCGTCTACAAAGGGGAAAAGCCTAGCGATCTC 2148
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        762 GGGATTGGTGGTGTACCGTTCCCCGTTCTCGTCTACAAAGGGGAAAAGCCTAGCGATCTC 703

Qy       2149 CTCTTCCGAATAGGGGCTAGCCGATTCGTTCCAAACGTAGTCCCGCGTTTTGGAGTAGAC 2208
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        702 CTCTTCCGAATAGGGGCTAGCCGATTCGTTCCAAACGTAGTCCCGCGTTTTGGAGTAGAC 643

Qy       2209 GAGGATCATGTCCTTTTGCGATCCGAAGGCCTTACGGGAAAAGTTTTTGGGATTTGAAGC 2268
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        642 GAGGATCATGTCCTTTTGCGATCCGAAGGCCTTACGGGAAAAGTTTTTGGGATTTGAAGC 583

Qy       2269 GATGCGGGCGATATGGTTAACGAAGTTTCGCCGGCCAAAGACCTCATCAAGGATGAGCTT 2328
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        582 GATGCGGGCGATATGGTTAACGAAGTTTCGCCGGCCAAAGACCTCATCAAGGATGAGCTT 523

Qy       2329 CACCTCGAACCCGTATTTCTCGTCTATGTGAACGAAGATCAGTCCTGAGTCCGCCATCAG 2388
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        522 CACCTCGAACCCGTATTTCTCGTCTATGTGAACGAAGATCAGTCCTGAGTCCGCCATCAG 463

Qy       2389 CTCCCTGAGAAGTATCAAGCGCTCCCTCAGGAACTCCACAAACTGAGGACCATCGAGGGT 2448
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        462 CTCCCTGAGAAGTATCAAGCGCTCCCTCAGGAACTCCACAAACTGAGGACCATCGAGGGT 403

Qy       2449 GTCATCGTAGCCCAACTGACCGTTTTTGGGCTGGCTGACGGTAGCAACGCGATCTGTTTC 2508
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        402 GTCATCGTAGCCCAACTGACCGTTTTTGGGCTGGCTGACGGTAGCAACGCGATCTGTTTC 343

Qy       2509 ATCGCCGCCAACGAGAAACTGCTGGCCGGTTCCATAAGGCGGGTCAATATAGACCAACTG 2568
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        342 ATCGCCGCCAACGAGAAACTGCTGGCCGGTTCCATAAGGCGGGTCAATATAGACCAACTG 283

Qy       2569 GACCTTCCCCGCATACCCACCAGGCTCCCGGAGCATCCACCGGAGAACCTGACCGTTTTC 2628
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        282 GACCTTCCCCGCATACCCACCAGGCTCCCGGAGCATCCACCGGAGAACCTGACCGTTTTC 223

Qy       2629 CCCCAAAAAGTAGGTGCCAATAGGATCAATCTCAAAAAGGGGGGCATTTCCCCCTAGGAA 2688
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        222 CCCCAAAAAGTAGGTGCCAATAGGATCAATCTCAAAAAGGGGGGCATTTCCCCCTAGGAA 163

Qy       2689 GAGGAGGGTTTCTTTTCGCAAAACAAGTTGTGGGGTGGGCTGATCAAGAATCTCCTTCTC 2748
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        162 GAGGAGGGTTTCTTTTCGCAAAACAAGTTGTGGGGTGGGCTGATCAAGAATCTCCTTCTC 103

Qy       2749 ATCGCGTTTTCCGGGGTAGACCAACCTAAAGGGCGAAGGTTCCGAGGTTTTCGAGGCTTT 2808
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        102 ATCGCGTTTTCCGGGGTAGACCAACCTAAAGGGCGAAGGTTCCGAGGTTTTCGAGGCTTT 43

Qy       2809 CAAGGGGGCTTTTCGGGTCAAACCAGGGTAGCTACGGCTCAT 2850
             |||||||||||||||||||||||||||||||||||||||||||
Db          42 CAAGGGGGCTTTTCGGGTCAAACCAGGGTAGCTACGGCTCAT 1
```

```
RESULT 3
AAD04666/c
ID   AAD04666 standard; DNA; 1026 BP.
XX
AC   AAD04666;
XX
DT   04-JUL-2001  (first entry)
XX
DE   Thermus replication protein, RepT encoding DNA from pTsp45S plasmid.
XX
KW   Replication protein; RepT; partition protein; ParA; pTsp45S plasmid;
KW   kanamycin-resistance gene; thermophilic transformation; Ori;
KW   replication origin; ds.
XX
OS   Thermus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .1026
FT                   /*tag= a
FT                   /product= "Replication protein, RepT"
FT                   /codon= (Seq:GTG, aa:Met)
XX
PN   US6207377-B1.
XX
PD   27-MAR-2001.
XX
PF   14-AUG-1998;   98US-00134246.
XX
PR   14-AUG-1998;   98US-00134246.
XX
PA   (NEWE ) NEW ENGLAND BIOLABS INC.
XX
PI   Wayne J,  Xu S;
XX
DR   WPI; 2001-298939/31.
DR   P-PSDB; AAE01000.
XX
PT   Cloning Thermus species (Ts) plasmid genes comprises transforming
```

```
PT   Escherichia coli with cloned recombinant plasmid containing Ts and E.coli
PT   origins of replication, isolating cloned recombinant plasmid from E.coli
PT   and transforming Ts cell.
XX
PS   Example I; Fig 1; 32pp; English.
XX
CC   The present DNA sequence encodes Thermus sp. replication protein, RepT
CC   which is obtained from pTsp45S plasmid. The replication protein, RepT is
CC   needed for thermophilic plasmid replication. The invention relates to
CC   Thermus sp. replication protein RepT, partition protein ParA and their
CC   corresponding DNA molecules which relates to recombinant DNA molecules
CC   encoding plasmid DNA replication origins in Thermus, as well as to
CC   shuttle vectors which contain the same. The invention also relates to
CC   method useful for cloning Thermus sp. plasmid genes which comprises
CC   inserting plasmid DNA comprising a Thermus sp. origin of replication
CC   (Ori) into a recombinant plasmid comprising a thermostable kanamycin-
CC   resistance gene and an Escherichia coli Ori, to produce a cloned
CC   recombinant plasmid. This cloned recombinant plasmid is transformed with
CC   an E. coli. host cell, and E. coli. host cell cultured for the expression
CC   of cloned recombinant plasmid. The cloned recombinant plasmid isolated
CC   from E. coli host cell is then transformed with Thermus sp. host cell and
CC   Thermus sp. host cell is cultured. Thus Thermus sp. plasmid genes are
CC   cloned. These plasmid DNAs are used for thermophilic transformation
XX
SQ   Sequence 1026 BP; 221 A; 323 C; 285 G; 197 T; 0 U; 0 Other;

  Query Match            17.5%;  Score 1026;  DB 5;  Length 1026;
  Best Local Similarity  100.0%;  Pred. No. 5.1e-273;
  Matches 1026;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3132 TCAGGAGGCCTCCTTGAGGGTCTTCACTAGGTACGCTCCCGGTCGTGTCAGATGGCCATC 3191
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1026 TCAGGAGGCCTCCTTGAGGGTCTTCACTAGGTACGCTCCCGGTCGTGTCAGATGGCCATC 967

Qy       3192 CGTGTAATCTCGGATAACCCGCATTAGGACGGCAAATAGATATTGCGCGGGGAGTTCACC 3251
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        966 CGTGTAATCTCGGATAACCCGCATTAGGACGGCAAATAGATATTGCGCGGGGAGTTCACC 907

Qy       3252 CCTGGCCACAGCCCACAGCAAGCCTGCATAGAAACGTCTTGAACGACGGTCATCTAGGAG 3311
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        906 CCTGGCCACAGCCCACAGCAAGCCTGCATAGAAACGTCTTGAACGACGGTCATCTAGGAG 847

Qy       3312 ATCGGCAATGTACGTAGCAATGAGGGTGATAAGGGCCGGGAGTTTGGAACGCTCCACCTC 3371
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        846 ATCGGCAATGTACGTAGCAATGAGGGTGATAAGGGCCGGGAGTTTGGAACGCTCCACCTC 787

Qy       3372 GGGGAGGACCAGGATGAGGCCCAGGTCAACGGCCACGGTCTTGGTGTTGGGCATCACCCT 3431
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        786 GGGGAGGACCAGGATGAGGCCCAGGTCAACGGCCACGGTCTTGGTGTTGGGCATCACCCT 727

Qy       3432 TTTCCCCTGAGCCCAGAGGACCAGCACGTCCAGGGTGGGGCGGATTCCGTGGTCCTGGTA 3491
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        726 TTTCCCCTGAGCCCAGAGGACCAGCACGTCCAGGGTGGGGCGGATTCCGTGGTCCTGGTA 667

Qy       3492 GGCCTTGACCCAGTTGAAGGAGAGCACGCCGTTGGCCATGTCTAGGGCGAGGTTCCTCCA 3551
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        666 GGCCTTGACCCAGTTGAAGGAGAGCACGCCGTTGGCCATGTCTAGGGCGAGGTTCCTCCA 607

Qy       3552 GGGGTAGATGTAGTCGTCCAGGGTGAGCCTGGCTTTCCCTGGCCTCAGCCGGACGGCCCA 3611
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        606 GGGGTAGATGTAGTCGTCCAGGGTGAGCCTGGCTTTCCCTGGCCTCAGCCGGACGGCCCA 547

Qy       3612 AAGGGTGCCGATGGCCCGGCGCTCCCCGTTGACGGTTTGGTGAAGGACGTCGGTGGCCAC 3671
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        546 AAGGGTGCCGATGGCCCGGCGCTCCCCGTTGACGGTTTGGTGAAGGACGTCGGTGGCCAC 487

Qy       3672 CAGGCCCTTTTTCTCAAGGACCTTCTTCCAGGCGTGGACGGTCTGCCGGGTGACCCCCAG 3731
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        486 CAGGCCCTTTTTCTCAAGGACCTTCTTCCAGGCGTGGACGGTCTGCCGGGTGACCCCCAG 427

Qy       3732 GTTGAGGGCCAACATCTCCAGGGGGACCATGAAGACGACCGTCCCCACCTTCCTCCAGAG 3791
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        426 GTTGAGGGCCAACATCTCCAGGGGGACCATGAAGACGACCGTCCCCACCTTCCTCCAGAG 367

Qy       3792 CTCCCGGTTGCCGTAGGGGATGGTGGAGCGGGCAATCTCCTGGAGGAGTTCCAGAAGCTT 3851
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        366 CTCCCGGTTGCCGTAGGGGATGGTGGAGCGGGCAATCTCCTGGAGGAGTTCCAGAAGCTT 307

Qy       3852 CTGAGGCCCGTCCTGGACGGCTTGACATACATTCCCAACGGGGGGTTCAGCTGGTGGTAT 3911
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        306 CTGAGGCCCGTCCTGGACGGCTTGACATACATTCCCAACGGGGGGTTCAGCTGGTGGTAT 247

Qy       3912 AGGGGGGGCCCCAACAGAGGAAAGGAAAGCCTCATGGTCTTTCTCGGGTAAAGCAATCCT 3971
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        246 AGGGGGGGCCCCAACAGAGGAAAGGAAAGCCTCATGGTCTTTCTCGGGTAAAGCAATCCT 187

Qy       3972 AAGTGCCTCTTTTTGGTATGTAAAGCCCTTCGCGAGGCGATTTTCGGCACCTCCATCTGG 4031
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        186 AAGTGCCTCTTTTTGGTATGTAAAGCCCTTCGCGAGGCGATTTTCGGCACCTCCATCTGG 127

Qy       4032 AGGGGGGTCGGTGGCCAAGAAGAAGTCCTCTGACCCCCTATCTGACCCCCTAGTGGCATC 4091
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        126 AGGGGGGTCGGTGGCCAAGAAGAAGTCCTCTGACCCCCTATCTGACCCCCTAGTGGCATC 67

Qy       4092 GGTGTTGTCGTGGGTTTCCTCTAAAGCCTCGTAAAGCTCTTCAAAGAAGGTTTTTTCGTT 4151
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          66 GGTGTTGTCGTGGGTTTCCTCTAAAGCCTCGTAAAGCTCTTCAAAGAAGGTTTTTTCGTT 7

Qy       4152 CTTCAC 4157
             ||||||
Db           6 CTTCAC 1
```

```
RESULT 4
AAX01849
ID   AAX01849 standard; DNA; 999 BP.
XX
AC   AAX01849;
XX
DT   13-APR-1999  (first entry)
XX
DE   Thermus sp. Tsp45I genomic DNA.
XX
KW   Thermostable enzyme; Tsp45I; restriction endonuclease; cleavage; ss.
XX
OS   Thermus sp.
XX
PN   US5866422-A.
XX
PD   02-FEB-1999.
XX
PF   29-OCT-1997;   97US-00960756.
XX
PR   29-OCT-1997;   97US-00960756.
XX
PA   (NEWE ) NEW ENGLAND BIOLABS INC.
XX
PI   Xu S,  Wayne J;
XX
DR   WPI; 1999-141942/12.
DR   P-PSDB; AAW92373.
XX
PT   DNA coding for Tsp45I restriction endonuclease and methylase from Thermus
PT   species YS45 - useful for recombinant production of thermostable enzyme
PT   in Escherichia coli.
XX
PS   Disclosure; Fig 2A-B; 22pp; English.
XX
CC   This sequence represents a novel restriction endonuclease gene, Tsp45I
CC   which has been isolated from Thermus sp. YS45. This DNA is useful for
CC   producing a thermostable endonuclease which can be used in a variety of
CC   assays, such as PCR, where high temperatures cannot be avoided. Tsp45I
CC   cleaves a double-stranded DNA molecule before the first G residue of the
CC   sequences 5'-GTGAC-3' and 5'-GTCAC-3'. The endonuclease creates sticky
CC   ends in the form of 5 nucleotide, single-stranded 5' overhangs. The
CC   production of recombinant Tsp45I in E. coli is simpler and produces 3
```

```
CC   multiply 105 units per gram of wet cell, about a ten-fold increase over
CC   that prepared from YS45
XX
SQ   Sequence 999 BP; 310 A; 219 C; 299 G; 171 T; 0 U; 0 Other;

  Query Match             17.1%;  Score 999;  DB 2;  Length 999;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-265;
  Matches  999;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        614 ATGCAACAGATGGCCGAGTGGAACGTGTGGACACAGAGAAGCGTTGAGCTTCTGGAGAAG 673
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCAACAGATGGCCGAGTGGAACGTGTGGACACAGAGAAGCGTTGAGCTTCTGGAGAAG 60

Qy        674 GGGTATTTGGATAAACTACTGCAGGTCTATAAAGGGGAAAGTGGCTCTTCGAGGTCAGTA 733
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGGTATTTGGATAAACTACTGCAGGTCTATAAAGGGGAAAGTGGCTCTTCGAGGTCAGTA 120

Qy        734 CCAGAGGAGGTAGAGGAAAAACTTCGCGAGGCCTACAAGGCATACGAGGGGAGGCAGGAT 793
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCAGAGGAGGTAGAGGAAAAACTTCGCGAGGCCTACAAGGCATACGAGGGGAGGCAGGAT 180

Qy        794 AGTCCGGAGGCAGAAACGAAACTCGTGGAAGCCGTGCTAAATGCCAGAAAAAAGGTCGAG 853
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGTCCGGAGGCAGAAACGAAACTCGTGGAAGCCGTGCTAAATGCCAGAAAAAAGGTCGAG 240

Qy        854 CGGTCCCCCTTCAATCACCCCTACCTGCCTTTGGTCTACTACCTGGTTTCGGAAAAAGCA 913
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CGGTCCCCCTTCAATCACCCCTACCTGCCTTTGGTCTACTACCTGGTTTCGGAAAAAGCA 300

Qy        914 GAAAAAGCGAACAAGGCCCTTGAGGAGGCATTGCAGGAGGTTGCCTCAAAGCACCCAGAA 973
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAAAAAGCGAACAAGGCCCTTGAGGAGGCATTGCAGGAGGTTGCCTCAAAGCACCCAGAA 360

Qy        974 ACCATCCGCGTCCTGGCCAAGGAAGCGCAAAGAAGAGGCGTAGAAGCCTTGATCCAAAGG 1033
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ACCATCCGCGTCCTGGCCAAGGAAGCGCAAAGAAGAGGCGTAGAAGCCTTGATCCAAAGG 420

Qy       1034 CTCAAGGAGCCTCCCGAAATAAATCGGCAGATAGGGCCGATGTTCAAAAGGTGGTACAAA 1093
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTCAAGGAGCCTCCCGAAATAAATCGGCAGATAGGGCCGATGTTCAAAAGGTGGTACAAA 480

Qy       1094 GAAGAGCTAAAGGGGAAAATAGAAGAGAGGCTTCCAGGCCCTACCAAACCAAAGATTGTG 1153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GAAGAGCTAAAGGGGAAAATAGAAGAGAGGCTTCCAGGCCCTACCAAACCAAAGATTGTG 540

Qy       1154 GTAGTATCCCCTGAAAAAAGTAAACCGGAGCAAGCACCCCTTATTGCGGAGAGAGAAGCG 1213
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTAGTATCCCCTGAAAAAAGTAAACCGGAGCAAGCACCCCTTATTGCGGAGAGAGAAGCG 600

Qy       1214 GGCATCATCATATACACGGGATCGGATGAAGCTTTGAAAGATGCCGCCAAGGAAAACCTG 1273
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCATCATCATATACACGGGATCGGATGAAGCTTTGAAAGATGCCGCCAAGGAAAACCTG 660

Qy       1274 GGCCTTGGCGAGGAAGCAGAACTAGGCACCAAGGGCGTAGATTTCTACGTGGTCATCCGG 1333
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GGCCTTGGCGAGGAAGCAGAACTAGGCACCAAGGGCGTAGATTTCTACGTGGTCATCCGG 720

Qy       1334 CGTAGCCCTGAAGAGACATGGCACCTAACAGGAGAAGTGAAGTTTCAATCCGACTTTGGC 1393
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CGTAGCCCTGAAGAGACATGGCACCTAACAGGAGAAGTGAAGTTTCAATCCGACTTTGGC 780

Qy       1394 GGAAACCAAGACAACCAGAAACTAGTAGCAAAGGCTTCCATAAGGTTGGACCTTGAGAAG 1453
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GGAAACCAAGACAACCAGAAACTAGTAGCAAAGGCTTCCATAAGGTTGGACCTTGAGAAG 840

Qy       1454 AGGCACATAGGAATAGTGGTGGTGGACGGAATGCCTGTGGTGAGCAAGTTTCGTGGGTGG 1513
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AGGCACATAGGAATAGTGGTGGTGGACGGAATGCCTGTGGTGAGCAAGTTTCGTGGGTGG 900

Qy       1514 GCCGGACTGGGGAAAGAAACGATCGTTACATCCGTACTCCTCCTTCCAGACCTGATAGCG 1573
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GCCGGACTGGGGAAAGAAACGATCGTTACATCCGTACTCCTCCTTCCAGACCTGATAGCG 960

Qy       1574 GAGCTCTACCAAAAGGGTGAAGAAGCCCTGGGCCTCTAG 1612
              ||||||||||||||||||||||||||||||||||||||||
Db         961 GAGCTCTACCAAAAGGGTGAAGAAGCCCTGGGCCTCTAG 999


RESULT 5
AAT67772/c
ID   AAT67772 standard; DNA; 1032 BP.
XX
AC   AAT67772;
XX
DT   29-JUL-1997  (first entry)
XX
DE   H. pylori cytoplasmic protein ORF 24816915.aa.
XX
KW   Vaccine; prevention; treatment; infection; identification;
KW   binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW   duodenal ulcer disease; chronic gastritis; diagnosis; cytoplasmic; ds.
XX
OS   Helicobacter pylori.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .1032
FT                   /*tag=  a
XX
PN   WO9640893-A1.
XX
PD   19-DEC-1996.
XX
PF   06-JUN-1996;   96WO-US009122.
XX
PR   07-JUN-1995;   95US-00487032.
PR   01-APR-1996;   96US-00630405.
XX
PA   (ASTR ) ASTRA AB.
XX
PI   Smith D,  Berglindh OT,  Mellgaerd BL;
XX
DR   WPI; 1997-052306/05.
DR   P-PSDB; AAW20333.
XX
PT   Helicobacter pylori nucleic acid sequences and related polypeptide(s) -
PT   useful for vaccines to treat or prevent H. pylori infection, and to
PT   detect Helicobacter.
XX
PS   Claim 9; Page; 1481pp; English.
XX
CC   The present sequence encodes a Helicobacter pylori cytoplasmic protein.
CC   The protein may be used in a vaccine to prevent or treat H. pylori
CC   infection or to identify H. pylori polypeptide binding compounds, useful
CC   as potential H. pylori life cycle activators or inhibitors. The genomic
CC   sequence of H. pylori (ATCC 55679) was determined from overlapping
CC   contigs generated by mechanically shearing the bacterial DNA. The
CC   sequences were analysed for ORF of at least 180 nucleotides, and the
CC   predicted coding regions defined by computer evaluation. To identify
CC   likely H. pylori antigens for vaccine development, the amino acid
CC   sequences predicted from various ORF were analysed for significant
CC   homology to other known or exported membrane proteins. Having identified
CC   and determined the sequences of interest, particular regions can be
CC   isolated from H. pylori by PCR amplification for recombinant polypeptide
CC   production, e.g. in E. coli hosts. Note: This DNA sequence is not
CC   reproduced in the specification and has been derived from the related
CC   specification, WO9719098
XX
SQ   Sequence 1032 BP; 386 A; 166 C; 170 G; 310 T; 0 U; 0 Other;

  Query Match             3.2%;  Score 185.6;  DB 2;  Length 1032;
  Best Local Similarity   52.0%;  Pred. No. 3e-40;
  Matches  475;  Conservative    0;  Mismatches  424;  Indels   15;  Gaps    2;

Qy       1664 ACCCTTCTAGTGAAGGCTTTGACCGCCTCCCAGGAGGCATCTATGCCGATGGATCGCCGC 1723
              | | |  |  || | |||| ||  |||     |||   ||| ||||| ||  ||   ||
Db        971 AGCTTGTTTTTGCAAGCTTGGATAGCCAAATCGGAATTATCAATGCCTATAAATTTTCGT 912
```



RESULT 6
AAT77452/c
ID AAT77452 standard; DNA; 1032 BP.
XX
AC AAT77452;
XX
DT 11-AUG-1997 (first entry)
XX
DE H. pylori cytoplasmic protein ORF 24816915.aa.
XX
KW Chronic gastritis; duodenal ulcer disease; activator; inhibitor;
KW bacterial life cycle; vaccine; immunisation; detection; antisense;
KW inhibition; cytoplasmic; ds.
XX
OS Helicobacter pylori.
XX
FH Key Location/Qualifiers
FT CDS 1. .1032
FT /*tag= a
XX
PN WO9719098-A1.
XX
PD 29-MAY-1997.
XX
PF 15-NOV-1996; 96WO-US018542.
XX
PR 17-NOV-1995; 95US-00561469.
XX
PA (ASTR) ASTRA AB.
XX
PI Smith DH;
XX
DR WPI; 1997-298052/27.
DR P-PSDB; AAW24634.
XX
PT Helicobacter pylori nucleic acid sequences and related proteins - used
PT for diagnostics and therapeutics.
XX
PS Claim 1; Page 100; 235pp; English.
XX
CC The present sequence encodes a Helicobacter pylori cytoplasmic protein.
CC H. pylori has been strongly linked to chronic gastritis and duodenal
CC ulcer disease. The nucleic acid sequences of the invention are used to
CC evaluate compounds, especially activators or inhibitors of bacterial life
CC cycle, for the ability to bind an H. pylori nucleic acid sequence. The
CC nucleic acid sequences, and corresponding proteins, are also useful for
CC generating vaccines for immunising subjects against H. pylori or for use
CC in detecting the presence of Helicobacter species in a sample. Antisense
CC nucleic acid sequences of these sequences are used to inhibit expression
CC of a gene from Helicobacter species. H. pylori whole genomic DNA was
CC isolated and nebulised to a median size of 2000 bp. Purified DNA
CC fragments were blunt-ended and ligated to unique BstXI-linker adapters in
CC 100-1000 fold molar excess. These linkers are complementary to the BstXI-
CC cut pMPX vectors, while the overhang is not self-complementary. Therefore
CC the linkers will not concatemerise nor will the cut vector re-ligate
CC itself easily. The linker-adapter inserts were ligated to each of the 20
CC pMPX vectors to construct a series of shotgun subclone libraries. The
CC purified DNA samples were then sequenced. Note: The ORF/protein reference
CC number for this sequence was obtained from the related specification,
CC WO9640893
XX
SQ Sequence 1032 BP; 386 A; 166 C; 170 G; 310 T; 0 U; 0 Other;

Query Match 3.2%; Score 185.6; DB 2; Length 1032;
Best Local Similarity 52.0%; Pred. No. 3e-40;
Matches 475; Conservative 0; Mismatches 424; Indels 15; Gaps 2;

```
Db        735 --AACTTGTTCTAAGGCATAAATTTTTTTTCTAGGGTTATTATTGTTAGAATACTCAATC 678

Qy       1964 AGCCCTTGAGCGTCTAGCTCGTCAAGCTTCTCCGGGGGATAGCGCCAATGCCGTCCAGGA 2023

Db        677 AAACCTTCTTTATCCCAACGCTCAAGTGTGGCAATATCAGTGCGCCAATGCCGCCCTTTT 618

Qy       2024 GGGGGAAGTATTCCTCGCCAAGGCCTTCCGGTAGGGCCATCCTTGGTTTCTCCAGGAGCA 2083

Db        617 GGAGGTAGCATACCTTTAAATGCTTTAGAACATTCGCCACTTTCCACTTCTCCTGGAGCA 558

Qy       2084 TGCAGGGGATTGGTGGTGTACCGTTCCCCGTTCTCGTCTACAAAGGGGAAAAGCCTAGCG 2143

Db        557 TGTATTGGAACGGTAGTGTAACGCCTTTTATCTTTGTCAATTTTAGGGAATCGTTTTTCT 498

Qy       2144 ATCTCCTCTTCCGAATAGGGGCTAGCCGATTCGTTCCAAACGTAGTCCCGCGTTTTGGAG 2203

Db        497 AAATCTTGTGGCGTATAAGGGATCTTAGGTTCGTTAAAAATGGGATTTTTTCCTTTAGAG 438

Qy       2204 TAGACGAGGATCATGTCCTTTTGCGATCCGAAGGCCTTACGGGAAAAGTTTTTGGGATTT 2263

Db        437 TAAAATAAAATCATATCTTTTATGTTACCATAGCCTATCCTTTTAAAATTTTTAGGATTG 378

Qy       2264 GAAGCGATGCGGGCGATATGGTTAACGAAGTTTCGCCGGCCAAAGACCTCATCAAGGATG 2323

Db        377 CACTTTATGCGTGTGATTTCATTTCTAAAATTTTGTATGCCAAATATTTCATCTAACATT 318

Qy       2324 AGCTTCACCTCGAACCCGTATTTCTCGTCTATGTGAACGAAGATCAGTCCTGAGTCCGCC 2383

Db        317 ACCTTGACATAATGCCCTATCTTGTAATCTGTATGCACATAGATAGAGCCTTGTTCTGAA 258

Qy       2384 ATCAGCTCCCTGAGAAGTATCAAGCGCTCCCTCAGGAACTCCACAAACTGAGGACCATCG 2443

Db        257 AGCAATTCTTTAAGCAATACCAGGCGTTGTTTTAAAAATTCCATAAAATCCATACCCACT 198

Qy       2444 AGGGTGTCATCGTAGCCCAACTGACCGTTTTTGGGCTGGCTGACGGTAGCAACGCGATCT 2503

Db        197 ACTTTATCACTATAAGCAATATCGCCATTCTTAGAATTGCTAATTGTGGTTGCTCTACCA 138

Qy       2504 GTTTCATCGCCGCCAACGAGAAACTGCTGGCCGGTTCCATAAGGCGGGTCAATATAGACC 2563

Db        137 TTT---------GTGATAGTAAAATGATTGTTTGTAGCAAATGGAGGGTCAATATATATC 87

Qy       2564 AACTGGACCTTCCC 2577

Db          86 AAATCTATTTTTCC 73
```

RESULT 7
AAD04667/c
ID AAD04667 standard; DNA; 180 BP.
XX
AC AAD04667;
XX
DT 04-JUL-2001 (first entry)
XX
DE Thermus promoter sequence upstream of replication protein repT gene.
XX
KW Replication protein; RepT; partition protein; ParA; pTsp45S plasmid;
KW kanamycin-resistance gene; thermophilic transformation; Ori;
KW replication origin; promoter; ds.
XX
OS Thermus sp.
XX

FH	Key	Location/Qualifiers
FT	promoter	1. .150
FT		/*tag= a
FT	-35_signal	27. .32
FT		/*tag= b
FT	-10_signal	50. .55
FT		/*tag= c
FT	-35_signal	85. .90
FT		/*tag= d
FT	-10_signal	109. .114
FT		/*tag= e
FT	RBS	139. .143
FT		/*tag= f
FT		/standard_name= "Shine-Dalgarno Sequence"

XX
PN US6207377-B1.
XX
PD 27-MAR-2001.
XX
PF 14-AUG-1998; 98US-00134246.
XX
PR 14-AUG-1998; 98US-00134246.
XX
PA (NEWE) NEW ENGLAND BIOLABS INC.
XX
PI Wayne J, Xu S;
XX
DR WPI; 2001-298939/31.
XX
PT Cloning Thermus species (Ts) plasmid genes comprises transforming
PT Escherichia coli with cloned recombinant plasmid containing Ts and E.coli
PT origins of replication, isolating cloned recombinant plasmid from E.coli
PT and transforming Ts cell.
XX
PS Example I; Fig 2; 32pp; English.
XX
CC The present DNA sequence is Thermus sp. promoter sequence upstream of
CC replication protein, repT gene. The invention relates to Thermus sp.
CC replication protein RepT, partition protein ParA and their corresponding
CC DNA molecules which relates to recombinant DNA molecules encoding plasmid
CC DNA replication origins in Thermus, as well as to shuttle vectors which
CC contain the same. The invention also relates to method useful for cloning
CC Thermus sp. plasmid genes which comprises inserting plasmid DNA
CC comprising a Thermus sp. origin of replication (Ori) into a recombinant
CC plasmid comprising a thermostable kanamycin-resistance gene and an
CC Escherichia coli Ori, to produce a cloned recombinant plasmid. This
CC cloned recombinant plasmid is transformed with an E. coli. host cell, and
CC E. coli. host cell cultured for the expression of cloned recombinant
CC plasmid. The cloned recombinant plasmid isolated from E. coli host cell
CC is then transformed with Thermus sp. host cell and Thermus sp. host cell
CC is cultured. Thus Thermus sp. plasmid genes are cloned. These plasmid
CC DNAs are used for thermophilic transformation
XX
SQ Sequence 180 BP; 48 A; 44 C; 56 G; 32 T; 0 U; 0 Other;

Query Match 3.1%; Score 180; DB 5; Length 180;
Best Local Similarity 100.0%; Pred. No. 4e-39;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       4128 CTCTTCAAAGAAGGTTTTTTCGTTCTTCACCCTCGGACCTCCTTGTCATCTGGAGCCCGA 4187

Db        180 CTCTTCAAAGAAGGTTTTTTCGTTCTTCACCCTCGGACCTCCTTGTCATCTGGAGCCCGA 121

Qy       4188 GGCGTTACCCTAGGTCTTGGGGGTGATCCGGGGCAACCGCCTCGGTTTCGCCTTTTTATG 4247

Db        120 GGCGTTACCCTAGGTCTTGGGGGTGATCCGGGGCAACCGCCTCGGTTTCGCCTTTTTATG 61

Qy       4248 GGTCCAAAATAACCGTCAGCCCAGCGGCTGGCAATCCCCCCTCCTAAAAGGCCGTTATAG 4307

Db          60 GGTCCAAAATAACCGTCAGCCCAGCGGCTGGCAATCCCCCCTCCTAAAAGGCCGTTATAG 1
```

RESULT 8
ADA71938/c
ID ADA71938 standard; DNA; 2000 BP.
XX
AC ADA71938;
XX
DT 20-NOV-2003 (first entry)
XX
DE Rice gene, SEQ ID 5263.
XX

KW Plant; bacterial infection; fungal infection; viral infection; rice;
KW gene; ds.
XX
OS Oryza sativa.
XX
PN WO2003000898-A1.
XX
PD 03-JAN-2003.
XX
PF 22-JUN-2001; 2001WO-IB001105.
XX
PR 22-JUN-2001; 2001WO-IB001105.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX
DR WPI; 2003-175290/17.
XX
PT Identifying at least one gene involved in plant resistance or response to
PT pathogenic infection for conferring resistance or tolerance to a plant to
PT bacterial, fungal or viral infection by determining or detecting plant
PT gene expression.
XX
PS Claim 27; SEQ ID NO 5263; 899pp; English.
XX
CC The present invention relates to a method (M1) for identifying genes
CC involved in plant resistance or response to pathogenic infection. M1
CC comprises identifying a gene whose expression is significantly altered in
CC the incompatible interaction of plant gene expression relative to
CC expression of the gene in an uninfected plant, in a mutant plant that
CC does not express a gene associated with response to pathogenic infection,
CC or in a corresponding incompatible or compatible interaction. (M1) is
CC useful for conferring resistance to resistance or tolerance to a plant to
CC bacterial, fungal or viral infection. The present sequence was used to
CC illustrate the invention.
XX
SQ Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;

Query Match 1.0%; Score 59; DB 7; Length 2000;
Best Local Similarity 8.4%; Pred. No. 5.1e-05;
Matches 61; Conservative 339; Mismatches 324; Indels 1; Gaps 1;

```
Qy       2924 GCCAAAGGCGGCCTACCTTTTGAGCCCGTATCTTCCCCTGACGTATAGACCTTCGGATCG 2983
Db        746 KSYRMWSGMGRMRWSAWRYCSRMKCAKTKYASSARWTKRAKRSYRYRRRWYWKRKGWTYR 687

Qy       2984 TCTCAGGGTGCACCCGAAGGATGTCTGCAAGCTCCTCGGGGGTCAGGTACACGGGCTTCA 3043
Db        686 YRYWRSCRMTRARMSKRRKWAGASMKSCWMYWRGARSMWYSKYSCSAKCCKKTRYMTSSY 627

Qy       3044 TCCTCATGACACAACCTTACCCCACAGAGGACAACACATGCAACTATGGGCAAAGTAGAC 3103
Db        626 MSTGMYGMYSSYKSMSWTSKMSYMGKMTCTMYTSMKGSTRRSKMGRWSGMSRMYMRWWKK 567

Qy       3104 AACGAGACCAAAAGCTTGGGCCACTCTCTCAGGAGGCCTCCTTGAGGGTCTTCACTAGGT 3163
Db        566 MRKRKYMRYMKWKCTWRRCMCYRWGYTMYTTSRSRMMYTGRYKARYTSKRRYMWYKYRKY 507

Qy       3164 ACGCTCCCGGTCGTGTCAGATGGCCATCCGTGTAATCTCGGATAACCCGCATTAGGACGG 3223
Db        506 CWYYYYGMYMKCSYMMRYGYCKACKKCCYAMCWKAAYSGMMMYWYRKYSKWMRMSTKYMW 447

Qy       3224 CAAATAGATATTGCGCGGGGAGTTCACCCCTGGCCACAGCCCACAGCAAGCCTGCATAGA 3283
Db        446 SMWYKKCRSMKYGAKGCYGCKMWTYCSYGYMKWYTYMGSYKYSRCYKYMRMYMYKGWMYM 387

Qy       3284 AACGTCTTGAACGACGGTCATCTAGGAGATCGGCAATGTACGTAGCAATGAGGGTGATAA 3343
Db        386 MYYSAYSSMMTWYYYYAKYWKYWYKRRGTMSWYGKSYKKKYCTWWCYMKCMRCYRWRKMM 327

Qy       3344 GGGCCGGGAGTTTGGAACGCTCCACCTCGGGGAGGACCAGGATGAGGCCCAGGTCAACGG 3403
Db        326 RKKTKYSKRCYCWRYATC-YWCCCYRKRGWYSRRSMMRTAGKWKMRSWSRWCRSYSWYKM 268

Qy       3404 CCACGGTCTTGGTGTTGGGCATCACCCTTTTCCCCTGAGCCCAGAGGACCAGCACGTCCA 3463
Db        267 YKKMWKKSYYMSYGWARSSGTWSRSAAKRTYKGYSTSRRAKMMRACRMYSACRRYSRTSY 208

Qy       3464 GGGTGGGGCGGATTCCGTGGTCCTGGTAGGCCTTGACCCAGTTGAAGGAGAGCACGCCGT 3523
Db        207 YCGCSYCGSSKWKYMSKSCSMRMTCSSWCSCCYTCYYGAMCWSCCMSMMYMGSCGCYTRG 148

Qy       3524 TGGCCATGTCTAGGGCGAGGTTCCTCCAGGGGTAGATGTAGTCGTCCAGGGTGAGCCTGG 3583
Db        147 WKWRSKYSMCCKKYCSCCTKYCSYTGYYRYCKWYKYSYYKCYYCYCYWYMSYMRYMMKCM 88

Qy       3584 CTTTCCCTGGCCTCAGCCGGACGGCCCAAAGGGTGCCGATGGCCCGGCGCTCCCCGTTGA 3643
Db         87 CSRSCSSWMSCAYCSTSSTSRWMSMYYAAKMGMCGSSGMYRMSKSCKMYSKYSSCKYTGS 28

Qy       3644 CGGTT 3648
Db         27 KKCTK 23
```

RESULT 9
ADA71938
ID ADA71938 standard; DNA; 2000 BP.
XX
AC ADA71938;
XX
DT 20-NOV-2003 (first entry)
XX
DE Rice gene, SEQ ID 5263.
XX
KW Plant; bacterial infection; fungal infection; viral infection; rice;
KW gene; ds.
XX
OS Oryza sativa.
XX
PN WO2003000898-A1.
XX
PD 03-JAN-2003.
XX
PF 22-JUN-2001; 2001WO-IB001105.
XX
PR 22-JUN-2001; 2001WO-IB001105.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX
DR WPI; 2003-175290/17.
XX
PT Identifying at least one gene involved in plant resistance or response to
PT pathogenic infection for conferring resistance or tolerance to a plant to
PT bacterial, fungal or viral infection by determining or detecting plant
PT gene expression.
XX
PS Claim 27; SEQ ID NO 5263; 899pp; English.
XX
CC The present invention relates to a method (M1) for identifying genes
CC involved in plant resistance or response to pathogenic infection. M1
CC comprises identifying a gene whose expression is significantly altered in
CC the incompatible interaction of plant gene expression relative to
CC expression of the gene in an uninfected plant, in a mutant plant that
CC does not express a gene associated with response to pathogenic infection,
CC or in a corresponding incompatible or compatible interaction. (M1) is
CC useful for conferring resistance to resistance or tolerance to a plant to
CC bacterial, fungal or viral infection. The present sequence was used to
CC illustrate the invention.
XX
SQ Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;

Query Match 1.0%; Score 57.6; DB 7; Length 2000;
Best Local Similarity 10.6%; Pred. No. 0.00012;
Matches 97; Conservative 394; Mismatches 418; Indels 5; Gaps 2;

```
Qy       621 AGATGGCCGAGTGGAACGTGTGGACACAGAGAAGCGTTGAGCTTCTGGAGAAGGGGTATT 680
Db        16 RGRRRMYMAGMMSCARMGSSRMSRKMGSMSKYRKCSSCGKCKMTTRRKSKWYSASSASGR 75

Qy       681 TGGATAAACTACTGCAGGTCTATAAAGGGGAAAGTGGCTCTTCGAGGTCAGTACCAGAGG 740
Db        76 TGSKWSSGSYSGKGMKKRYKRSKRWRGRGRRGMRRSRMRWMGRYRRCARSGRMAGGSGRM 135

Qy       741 AGGTAGAGGAAAAACTTCGCGAGGCCTACAAGGCATACGAGGGGAGGCAGGATAGTCCGG 800
Db       136 MGGKSRMSYWMWCYARGCGSCKRKKSKGGSWGKTCRRGARGGSGWSSGAKYKSGSMSKRM 195

Qy       801 AGGCAGAAACGAAACTCGTGGAAGCCGTGCTAAATGCCAGAAAAAAGGTCGAGCGGTCCC 860
Db       196 WMSSCGRSGCGRRSAYSRYYGTSRKYGTYKKMTYYSASRCMRAYMTTSYSWACSSYTWCR 255

Qy       861 CCTTCAATCACCCCTACCTGCCTTTGGTCTACTACCTGGTTTCGGAAAAAGCAGAAAAAG 920
Db       256 SKRRSMMWKMMRKMRWSRSYGWYSWSYKMWMCTAYKKSYYSRWCYMYRGGGWRGATRYWG 315

Qy       921 CGAACAAGGCCCTTGAGGAGGCATTGCAGGAGGTTGCCTCAAAGCACCCAGAAACCATCC 980
Db       316 RGYMSRMAMMYKKMYWYRGYKGMKRGWWAGRMMMRSMC---RWSKACYYMRWRMWRMTRR 372

Qy       981 GCGTCCTGGCCAAGGAAGCGCAAAGAAGAGGCGTAGAAGCCTTGATCCAAAGGCTCAAGG 1040
Db       373 RRWAKKSSRTSRRKKRKWCMRKRKYKRMRGYSRMRSCKRARWMKRCRSGRAWKMGCRGCM 432

Qy      1041 AGCCTCCCGAAATAAATCGGCAGATAGGGCCGATGTTCAAAAGGTGGTACAAAGAAGAGC 1100
Db       433 TCRMKSYGMMRWKSWKRMASKYKWMSRMYRWRKKKCSRTTMWGKTRGGMMGTMGRCRYKK 492

Qy      1101 TAAAGGGGAAAATAGAAGAGAGGCTTCCAGGCCCTACCAAACCAAAGATTGTGGTAGTAT 1160
Db       493 RSGMKRKCRRRRWGRMYRMRWKRYYMSARYTMRYCARKKYSYSAARKARCWYRGKGYYWA 552

Qy      1161 CCCCTGAAAAAAGTAAACCGGAGCAAGCACCCCTTATTGCGGAGAGAGAAGCGGGCATCA 1220
Db       553 GMWMKRYKRMYMYKMMWWYKRKYSKCSWYCKMSYYASCMKSARKAGAKMCKRSKMSAWSK 612

Qy      1221 TCATATACACGGGATCGGATGAAGCTTTGAAAGATGCCGCCAAGGAAAACCTGGGCCTTG 1280
Db       613 SMRSSRKCRKCASKRSSAKRYAMMGGMTSGSRMSRWKSYTCYWRKWGSMKSTCTWMYYMS 672

Qy      1281 GCGAGGAAGCAGAACTAGGCACCAAGGGCGTAGATTTCTACGTGGTCATCCGGCGTAGCC 1340
Db       673 KYTYAKYGSYWRYRYRAWCMYMWRWYYYRYRSYMTYMAWYTSSTRMAMTGMKYSGRYWTS 732

Qy      1341 CTGAAGAGACATGGCACCTAACAGGAGAAGTGAAGTTTCAATCCGACTTTGGCGGAAACC 1400
Db       733 WYKYCKCSWKYRSMWYYWSWWWAKTWKMWRRYATRMMWMWYRYSMKWYTWCTMWGYWWY- 791

Qy      1401 AAGACAACCAGAAACTAGTAGCAAAGGCTTCCATAAGGTTGGACCTTGAGAAGAGGCACA 1460
Db       792 -WWRTYMKMRYMWYKCTKTYWYWSATYWTGTWAAWWMAKTKMRMGMTGAKTRGRARKARY 850

Qy      1461 TAGGAATAGTGGTGGTGGACGGAATGCCTGTGGTGAGCAAGTTTCGTGGGTGGGCCGGAC 1520
Db       851 WWKWATWCATKRWMTKGKGAKWAWTWMAKAWRKYYWSWMRAWYYYYKTRRTRYKTCWWKA 910

Qy      1521 TGGGGAAAGAAACG 1534
Db       911 RWGSWAYWRMWWKG 924
```

RESULT 10
ADA70342/c
ID ADA70342 standard; DNA; 2030 BP.

XX
AC ADA70342;
XX
DT 20-NOV-2003 (first entry)
XX
DE Rice gene, SEQ ID 3665.
XX
KW Plant; bacterial infection; fungal infection; viral infection; rice;
KW gene; ds.
XX
OS Oryza sativa.
XX
PN WO2003000898-A1.
XX
PD 03-JAN-2003.
XX
PF 22-JUN-2001; 2001WO-IB001105.
XX
PR 22-JUN-2001; 2001WO-IB001105.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX
DR WPI; 2003-175290/17.
XX
PT Identifying at least one gene involved in plant resistance or response to
PT pathogenic infection for conferring resistance or tolerance to a plant to
PT bacterial, fungal or viral infection by determining or detecting plant
PT gene expression.
XX
PS Claim 6; SEQ ID NO 3665; 899pp; English.
XX
CC The present invention relates to a method (M1) for identifying genes
CC involved in plant resistance or response to pathogenic infection. M1
CC comprises identifying a gene whose expression is significantly altered in
CC the incompatible interaction of plant gene expression relative to
CC expression of the gene in an uninfected plant, in a mutant plant that
CC does not express a gene associated with response to pathogenic infection,
CC or in a corresponding incompatible or compatible interaction. (M1) is
CC useful for conferring resistance to resistance or tolerance to a plant to
CC bacterial, fungal or viral infection. The present sequence was used to
CC illustrate the invention.
XX
SQ Sequence 2030 BP; 388 A; 659 C; 727 G; 246 T; 0 U; 10 Other;

Query Match 1.0%; Score 56.4; DB 7; Length 2030;
Best Local Similarity 46.7%; Pred. No. 0.00027;
Matches 214; Conservative 0; Mismatches 241; Indels 3; Gaps 1;

```
Qy      3348 CGGGAGTTTGGAACGCTCCACCTCGGGGAGGACCAGGATGAGGCCCAGGTCAACGGCCAC 3407
Db       942 CGCGCGCTTGGCGCGCTCGCACTCGCGGCGGAGCTTGCCCAGCGCGCGCGCGTCGCCGGC 883

Qy      3408 GGTCTTGGTGTTGGGCATCACCCTTTTCCCCTGAGCCCAGAGGACCAGCACGTCCAGGGT 3467
Db       882 GATGTCGCGCCCGTGCTTCCGGCGGATGACCTTGACGAAGTGGTCCATGAGGCGTTGGTC 823

Qy      3468 GGGG---CGGATTCCGTGGTCCTGGTAGGCCTTGACCCAGTTGAAGGAGAGCACGCCGTT 3524
Db       822 GAAGTCCTCGCCGCCGAGGTGGGTGTCGCCGTTGGTGGCAAGGACCTCGAACACGCCGTT 763

Qy      3525 GGCCATGTCTAGGGCGAGGTTCCTCCAGGGGTAGATGTAGTCGTCCAGGGTGAGCCTGGC 3584
Db       762 GTCGATGGCGAGGATGCTGACGTCGAACGTGCCGCCGCCGAGGTCGAAGACGAGGACGTT 703

Qy      3585 TTTCCCTGGCCTCAGCCGGACGGCCCAAAGGGTGCCGATGGCCCGGCGCTCCCCGTTGAC 3644
Db       702 CTTCTCGGCGCCCTTCTTGTCGATGCCGTAGGCGATGGCGGCGGCGGTCGGCTCGTTGAT 643

Qy      3645 GGTTTGGTGAAGGACGTCGGTGGCCACCAGGCCCTTTTTCTCAAGGACCTTCTTCCAGGC 3704
```

```
Db        642 GATGCGGTCGACGGTGAGCCCGGCGATGACGCCGGCGTCCTTGGTGGCCTGCCGCTGCGC 583

Qy       3705 GTGGACGGTCTGCCGGGTGACCCCCAGGTTGAGGGCCAACATCTCCAGGGGGACCATGAA 3764

Db        582 GTCGTTGAAGTAGGCCGGGACGGTGACGACGGCGCGCGTGACCTTCTCGCCGAGGTAGGC 523

Qy       3765 GACGACCGTCCCCACCTTCCTCCAGAGCTCCCGGTTGC 3802

Db        522 CTCGGCCGTCTCCTTCATCCGCGTGAGCACCATGGCGC 485


RESULT 11
ADC08176/c
ID   ADC08176 standard; DNA; 2031 BP.
XX
AC   ADC08176;
XX
DT   18-DEC-2003  (first entry)
XX
DE   Rice DNA sequence Seq ID443 related to grain filling.
XX
KW   plant biotechnology; carbohydrate synthesis; carbohydrate metabolism;
KW   carbohydrate degradation; carbohydrate; plant grain; grain filling; corn;
KW   tomato; banana; canola; cotton; peanut; sorghum; tobacco; sugarbeet;
KW   wheat; rice; protein; oil; starch; fibre; moisture content; cereal grain;
KW   gene; ds; plant.
XX
OS   Oryza sativa.
XX
PN   WO2003000905-A2.
XX
PD   03-JAN-2003.
XX
PF   21-JUN-2002; 2002WO-IB002450.
XX
PR   22-JUN-2001; 2001US-0300112P.
PR   26-SEP-2001; 2001US-0325277P.
PR   20-DEC-2001; 2001US-0342327P.
XX
PA   (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
PI   Zhu T,  Cheng W,  Briggs S,  Cooper B,  Goff SA,  Moughamer T;
PI   Glazebrook J,  Katagiri F,  Kreps J,  Provart N,  Ricke D;
XX
DR   WPI; 2003-229341/22.
DR   P-PSDB; ADC08177.
XX
PT   New plant genes encoding polypeptides having an activity involved in or
PT   associated with the synthesis, metabolism or degradation of carbohydrates
PT   in the plant grain useful in generating plants having improved
PT   nutritional properties.
XX
PS   Claim 35; SEQ ID NO 443; 130pp; English.
XX
CC   This invention, in the area of plant biotechnology, relates to novel
CC   polynucleotides comprising a nucleotide sequence encoding a protein which
CC   is involved in or associated with the synthesis, metabolism or
CC   degradation of carbohydrates in the plant grain and the expression of
CC   which is up-regulated during grain filling. The plant is selected from
CC   corn, tomato, banana, canola, cotton, peanut, sorghum, tobacco,
CC   sugarbeet, wheat, and rice. The invention may be useful for the
CC   improvement of protein, oil, starch, fibre and moisture content of the
CC   cereal grains. In addition, carbohydrate levels may be modified to a more
CC   desirable level using the present invention. The present sequence is a
CC   DNA sequence encoding a rice protein of the invention. Note: The sequence
CC   data for this patent did not form part of the printed specification, but
CC   was obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/publishedpct_sequences.
XX
SQ   Sequence 2031 BP; 388 A; 659 C; 728 G; 246 T; 0 U; 10 Other;

  Query Match             1.0%;  Score 56.4;  DB 9;  Length 2031;
  Best Local Similarity   46.7%;  Pred. No. 0.00027;
  Matches  214;  Conservative    0;  Mismatches  241;  Indels    3;  Gaps    1;

Qy       3348 CGGGAGTTTGGAACGCTCCACCTCGGGGAGGACCAGGATGAGGCCCAGGTCAACGGCCAC 3407

Db        942 CGCGCGCTTGGCGCGCTCGCACTCGCGGCGGAGCTTGCCCAGCGCGCGCGCGTCGCCGGC 883

Qy       3408 GGTCTTGGTGTTGGGCATCACCCTTTTCCCCTGAGCCCAGAGGACCAGCACGTCCAGGGT 3467

Db        882 GATGTCGCGCCCGTGCTTCCGGCGGATGACCTTGACGAAGTGGTCCATGAGGCGTTGGTC 823

Qy       3468 GGGG---CGGATTCCGTGGTCCTGGTAGGCCTTGACCCAGTTGAAGGAGAGCACGCCGTT 3524

Db        822 GAAGTCCTCGCCGCCGAGGTGGGTGTCGCCGTTGGTGGCAAGGACCTCGAACACGCCGTT 763

Qy       3525 GGCCATGTCTAGGGCGAGGTTCCTCCAGGGGTAGATGTAGTCGTCCAGGGTGAGCCTGGC 3584

Db        762 GTCGATGGCGAGGATGCTGACGTCGAACGTGCCGCCGCCGAGGTCGAAGACGAGGACGTT 703

Qy       3585 TTTCCCTGGCCTCAGCCGGACGGCCCAAAGGGTGCCGATGGCCCGGCGCTCCCCGTTGAC 3644

Db        702 CTTCTCGGCGCCCTTCTTGTCGATGCCGTAGGCGATGGCGGCGGCGGTCGGCTCGTTGAT 643

Qy       3645 GGTTTGGTGAAGGACGTCGGTGGCCACCAGGCCCTTTTTCTCAAGGACCTTCTTCCAGGC 3704

Db        642 GATGCGGTCGACGGTGAGCCCGGCGATGACGCCGGCGTCCTTGGTGGCCTGCCGCTGCGC 583

Qy       3705 GTGGACGGTCTGCCGGGTGACCCCCAGGTTGAGGGCCAACATCTCCAGGGGGACCATGAA 3764

Db        582 GTCGTTGAAGTAGGCCGGGACGGTGACGACGGCGCGCGTGACCTTCTCGCCGAGGTAGGC 523

Qy       3765 GACGACCGTCCCCACCTTCCTCCAGAGCTCCCGGTTGC 3802

Db        522 CTCGGCCGTCTCCTTCATCCGCGTGAGCACCATGGCGC 485


RESULT 12
ADA48615/c
ID   ADA48615 standard; DNA; 2028 BP.
XX
AC   ADA48615;
XX
DT   20-NOV-2003  (first entry)
XX
DE   Rice gene conferring disease resistance in plants.
XX
KW   disease resistance; pathogen tolerance; plant pathogen; ds; gene; plant.
XX
OS   Oryza sativa.
XX
PN   WO2003000906-A2.
XX
PD   03-JAN-2003.
XX
PF   21-JUN-2002; 2002WO-IB002453.
XX
PR   22-JUN-2001; 2001US-0300112P.
PR   26-SEP-2001; 2001US-0352277P.
PR   22-MAR-2002; 2002US-0366535P.
XX
PA   (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
PI   Glazebrook J,  Briggs S,  Cooper B,  Goff SA,  Moughamer T;
PI   Katagiri F,  Kreps J,  Provart N,  Ricke D,  Zhu T;
XX
DR   WPI; 2003-184052/18.
DR   P-PSDB; ADA48616.
XX
PT   New polynucleotide comprising a plant nucleotide sequence having an open
PT   reading frame that encodes a polypeptide associated with disease
PT   resistance, useful for conferring resistance or tolerance to a plant
PT   pathogen.
XX
PS   Claim 1; SEQ ID NO 685; 299pp; English.
```

XX
CC The invention relates to a novel isolated polynucleotide comprising a
CC plant nucleotide sequence having an open reading frame that encodes a
CC polypeptide associated with disease resistance or its fragment having
CC substantially the same activity as the full-length polypeptide. The
CC polynucleotide of the invention is useful for conferring resistance or
CC tolerance to a plant pathogen. The present sequence represents a gene
CC conferring disease resistance used in the invention.
XX
SQ Sequence 2028 BP; 390 A; 662 C; 729 G; 246 T; 0 U; 1 Other;

Query Match 0.9%; Score 55.4; DB 8; Length 2028;
Best Local Similarity 46.5%; Pred. No. 0.00051;
Matches 213; Conservative 0; Mismatches 242; Indels 3; Gaps 1;

Qy 3348 CGGGAGTTTGGAACGCTCCACCTCGGGGAGGACCAGGATGAGGCCCAGGTCAACGGCCAC 3407
Db 942 CGCGCGCTTGGCGCGCTCGCACTCGCGGCGGAGCTTGCCCAGCGCGCGCGCGTCGCCGGC 883

Qy 3408 GGTCTTGGTGTTGGGCATCACCCTTTTCCCCTGAGCCCAGAGGACCAGCACGTCCAGGGT 3467
Db 882 GATGTCGCGCCCGTGCTTCCGGCGGATGACCTTGACGAAGTGGTCCATGAGGCGTTGGTC 823

Qy 3468 GGGG---CGGATTCCGTGGTCCTGGTAGGCCTTGACCCAGTTGAAGGAGAGCACGCCGTT 3524
Db 822 GAAGTCCTCGCCGCCGAGGTGGGTGTCGCCGTTGGTGGCAAGGACCTCGAACACGCCGTT 763

Qy 3525 GGCCATGTCTAGGGCGAGGTTCCTCCAGGGGTAGATGTAGTCGTCCAGGGTGAGCCTGGC 3584
Db 762 GTCGATGGCGAGGATGCTGACGTCGAACGTGCCGCCGCCGAGGTCGAAGACGAGGACGTT 703

Qy 3585 TTTCCCTGGCCTCAGCCGGACGGCCCAAAGGGTGCCGATGGCCCGGCGCTCCCCGTTGAC 3644
Db 702 CTTCTCNGCGCCCTTCTTGTCGATGCCGTAGGCGATGGCGGCGGCGGTCGGCTCGTTGAT 643

Qy 3645 GGTTTGGTGAAGGACGTCGGTGGCCACCAGGCCCTTTTTCTCAAGGACCTTCTTCCAGGC 3704
Db 642 GATGCGGTCGACGGTGAGCCCGGCGATGACGCCGGCGTCCTTGGTGGCCTGCCGCTGCGC 583

Qy 3705 GTGGACGGTCTGCCGGGTGACCCCCAGGTTGAGGGCCAACATCTCCAGGGGGACCATGAA 3764
Db 582 GTCGTTGAAGTAGGCCGGGACGGTGACGACGGCGCGCGTGACCTTCTCGCCGAGGTAGGC 523

Qy 3765 GACGACCGTCCCCACCTTCCTCCAGAGCTCCCGGTTGC 3802
Db 522 CTCGGCCGTCTCCTTCATCCGGGTGAGCACCATGGCGC 485

RESULT 13
ADC08212/c
ID ADC08212 standard; DNA; 2028 BP.
XX
AC ADC08212;
XX
DT 18-DEC-2003 (first entry)
XX
DE Rice DNA sequence Seq ID517 related to grain filling.
XX
KW plant biotechnology; carbohydrate synthesis; carbohydrate metabolism;
KW carbohydrate degradation; carbohydrate; plant grain; grain filling; corn;
KW tomato; banana; canola; cotton; peanut; sorghum; tobacco; sugarbeet;
KW wheat; rice; protein; oil; starch; fibre; moisture content; cereal grain;
KW gene; ds; plant.
XX
OS Oryza sativa.
XX
PN WO2003000905-A2.
XX
PD 03-JAN-2003.
XX
PF 21-JUN-2002; 2002WO-IB002450.
XX
PR 22-JUN-2001; 2001US-0300112P.
PR 26-SEP-2001; 2001US-0325277P.
PR 20-DEC-2001; 2001US-0342327P.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Zhu T, Cheng W, Briggs S, Cooper B, Goff SA, Moughamer T;
PI Glazebrook J, Katagiri F, Kreps J, Provart N, Ricke D;
XX
DR WPI; 2003-229341/22.
DR P-PSDB; ADC08213.
XX
PT New plant genes encoding polypeptides having an activity involved in or
PT associated with the synthesis, metabolism or degradation of carbohydrates
PT in the plant grain useful in generating plants having improved
PT nutritional properties.
XX
PS Claim 35; SEQ ID NO 517; 130pp; English.
XX
CC This invention, in the area of plant biotechnology, relates to novel
CC polynucleotides comprising a nucleotide sequence encoding a protein which
CC is involved in or associated with the synthesis, metabolism or
CC degradation of carbohydrates in the plant grain and the expression of
CC which is up-regulated during grain filling. The plant is selected from
CC corn, tomato, banana, canola, cotton, peanut, sorghum, tobacco,
CC sugarbeet, wheat, and rice. The invention may be useful for the
CC improvement of protein, oil, starch, fibre and moisture content of the
CC cereal grains. In addition, carbohydrate levels may be modified to a more
CC desirable level using the present invention. The present sequence is a
CC DNA sequence encoding a rice protein of the invention. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/publishedpct_sequences.
XX
SQ Sequence 2028 BP; 390 A; 662 C; 729 G; 246 T; 0 U; 1 Other;

Query Match 0.9%; Score 55.4; DB 9; Length 2028;
Best Local Similarity 46.5%; Pred. No. 0.00051;
Matches 213; Conservative 0; Mismatches 242; Indels 3; Gaps 1;

Qy 3348 CGGGAGTTTGGAACGCTCCACCTCGGGGAGGACCAGGATGAGGCCCAGGTCAACGGCCAC 3407
Db 942 CGCGCGCTTGGCGCGCTCGCACTCGCGGCGGAGCTTGCCCAGCGCGCGCGCGTCGCCGGC 883

Qy 3408 GGTCTTGGTGTTGGGCATCACCCTTTTCCCCTGAGCCCAGAGGACCAGCACGTCCAGGGT 3467
Db 882 GATGTCGCGCCCGTGCTTCCGGCGGATGACCTTGACGAAGTGGTCCATGAGGCGTTGGTC 823

Qy 3468 GGGG---CGGATTCCGTGGTCCTGGTAGGCCTTGACCCAGTTGAAGGAGAGCACGCCGTT 3524
Db 822 GAAGTCCTCGCCGCCGAGGTGGGTGTCGCCGTTGGTGGCAAGGACCTCGAACACGCCGTT 763

Qy 3525 GGCCATGTCTAGGGCGAGGTTCCTCCAGGGGTAGATGTAGTCGTCCAGGGTGAGCCTGGC 3584
Db 762 GTCGATGGCGAGGATGCTGACGTCGAACGTGCCGCCGCCGAGGTCGAAGACGAGGACGTT 703

Qy 3585 TTTCCCTGGCCTCAGCCGGACGGCCCAAAGGGTGCCGATGGCCCGGCGCTCCCCGTTGAC 3644
Db 702 CTTCTCNGCGCCCTTCTTGTCGATGCCGTAGGCGATGGCGGCGGCGGTCGGCTCGTTGAT 643

Qy 3645 GGTTTGGTGAAGGACGTCGGTGGCCACCAGGCCCTTTTTCTCAAGGACCTTCTTCCAGGC 3704
Db 642 GATGCGGTCGACGGTGAGCCCGGCGATGACGCCGGCGTCCTTGGTGGCCTGCCGCTGCGC 583

Qy 3705 GTGGACGGTCTGCCGGGTGACCCCCAGGTTGAGGGCCAACATCTCCAGGGGGACCATGAA 3764
Db 582 GTCGTTGAAGTAGGCCGGGACGGTGACGACGGCGCGCGTGACCTTCTCGCCGAGGTAGGC 523

Qy 3765 GACGACCGTCCCCACCTTCCTCCAGAGCTCCCGGTTGC 3802
Db 522 CTCGGCCGTCTCCTTCATCCGGGTGAGCACCATGGCGC 485

RESULT 14

ADA70341/c
ID ADA70341 standard; DNA; 2028 BP.
XX
AC ADA70341;
XX
DT 20-NOV-2003 (first entry)
XX
DE Rice gene, SEQ ID 3664.
XX
KW Plant; bacterial infection; fungal infection; viral infection; rice;
KW gene; ds.
XX
OS Oryza sativa.
XX
PN WO2003000898-A1.
XX
PD 03-JAN-2003.
XX
PF 22-JUN-2001; 2001WO-IB001105.
XX
PR 22-JUN-2001; 2001WO-IB001105.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX
DR WPI; 2003-175290/17.
XX
PT Identifying at least one gene involved in plant resistance or response to
PT pathogenic infection for conferring resistance or tolerance to a plant to
PT bacterial, fungal or viral infection by determining or detecting plant
PT gene expression.
XX
PS Claim 6; SEQ ID NO 3664; 899pp; English.
XX
CC The present invention relates to a method (M1) for identifying genes
CC involved in plant resistance or response to pathogenic infection. M1
CC comprises identifying a gene whose expression is significantly altered in
CC the incompatible interaction of plant gene expression relative to
CC expression of the gene in an uninfected plant, in a mutant plant that
CC does not express a gene associated with response to pathogenic infection,
CC or in a corresponding incompatible or compatible interaction. (M1) is
CC useful for conferring resistance to resistance or tolerance to a plant to
CC bacterial, fungal or viral infection. The present sequence was used to
CC illustrate the invention.
XX
SQ Sequence 2028 BP; 390 A; 662 C; 729 G; 246 T; 0 U; 1 Other;

Query Match 0.9%; Score 54.8; DB 7; Length 2028;
Best Local Similarity 46.5%; Pred. No. 0.00074;
Matches 213; Conservative 0; Mismatches 242; Indels 3; Gaps 1;

```
Qy       3348 CGGGAGTTTGGAACGCTCCACCTCGGGGAGGACCAGGATGAGGCCCAGGTCAACGGCCAC 3407
             || | | |||| ||||| |||| || ||| | |  || | | | || | |
Db        942 CGCGCGCTTGGCGCGCTCGCACTCGCGGCGGAGCTTGCCCAGCGCGCGCGCGTCGCCGGC 883

Qy       3408 GGTCTTGGTGTTGGGCATCACCCTTTTCCCCTGAGCCCAGAGGACCAGCACGTCCAGGGT 3467
             | | | |     | || ||   |  |  |||   |  || || |||  | |    ||
Db        882 GATGTCGCGCCCGTGCTTCCGGCGGATGACCTTGACGAAGTGGTCCATGAGGCGTTGGTC 823

Qy       3468 GGGG---CGGATTCCGTGGTCCTGGTAGGCCTTGACCCAGTTGAAGGAGAGCACGCCGTT 3524
             |  |     |   ||| |||    || | | |||        ||    || |||||||||
Db        822 GAAGTCCTCGCCGCCGAGGTGGGTGTCGCCGTTGGTGGCAAGGACCTCGAACACGCCGTT 763

Qy       3525 GGCCATGTCTAGGGCGAGGTTCCTCCAGGGGTAGATGTAGTCGTCCAGGGTGAGCCTGGC 3584
             | | ||| | |||  |  |     | | |  |  |  |  ||| | |  |||  |
Db        762 GTCGATGGCGAGGATGCTGACGTCGAACGTGCCGCCGCCGAGGTCGAAGACGAGGACGTT 703

Qy       3585 TTTCCCTGGCCTCAGCCGGACGGCCCAAAGGGTGCCGATGGCCCGGCGCTCCCCGTTGAC 3644
              ||| | |  | |  |  | ||   |    || |  |  |||   |  |  | ||||||
Db        702 CTTCTCSGCGCCCTTCTTGTCGATGCCGTAGGCGATGGCGGCGGCGGTCGGCTCGTTGAT 643

Qy       3645 GGTTTGGTGAAGGACGTCGGTGGCCACCAGGCCCTTTTTCTCAAGGACCTTCTTCCAGGC 3704
             | |  |||  | | |      ||| | | |||    | ||     | ||| |  |   ||
Db        642 GATGCGGTCGACGGTGAGCCCGGCGATGACGCCGGCGTCCTTGGTGGCCTGCCGCTGCGC 583

Qy       3705 GTGGACGGTCTGCCGGGTGACCCCCAGGTTGAGGGCCAACATCTCCAGGGGGACCATGAA 3764
             || |  |   |     | |||    | |  |  |  |  || | |  |  ||  |
Db        582 GTCGTTGAAGTAGGCCGGGACGGTGACGACGGCGCGCGTGACCTTCTCGCCGAGGTAGGC 523

Qy       3765 GACGACCGTCCCCACCTTCCTCCAGAGCTCCCGGTTGC 3802
                || ||||| ||  | |||    |||| ||  |  ||
Db        522 CTCGGCCGTCTCCTTCATCCGGGTGAGCACCATGGCGC 485
```

RESULT 15
ADA71232/c
ID ADA71232 standard; DNA; 2060 BP.
XX
AC ADA71232;
XX
DT 20-NOV-2003 (first entry)
XX
DE Rice gene, SEQ ID 4555.
XX
KW Plant; bacterial infection; fungal infection; viral infection; rice;
KW gene; ds.
XX
OS Oryza sativa.
XX
PN WO2003000898-A1.
XX
PD 03-JAN-2003.
XX
PF 22-JUN-2001; 2001WO-IB001105.
XX
PR 22-JUN-2001; 2001WO-IB001105.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX
DR WPI; 2003-175290/17.
XX
PT Identifying at least one gene involved in plant resistance or response to
PT pathogenic infection for conferring resistance or tolerance to a plant to
PT bacterial, fungal or viral infection by determining or detecting plant
PT gene expression.
XX
PS Claim 6; SEQ ID NO 4555; 899pp; English.
XX
CC The present invention relates to a method (M1) for identifying genes
CC involved in plant resistance or response to pathogenic infection. M1
CC comprises identifying a gene whose expression is significantly altered in
CC the incompatible interaction of plant gene expression relative to
CC expression of the gene in an uninfected plant, in a mutant plant that
CC does not express a gene associated with response to pathogenic infection,
CC or in a corresponding incompatible or compatible interaction. (M1) is
CC useful for conferring resistance to resistance or tolerance to a plant to
CC bacterial, fungal or viral infection. The present sequence was used to
CC illustrate the invention.
XX
SQ Sequence 2060 BP; 395 A; 670 C; 739 G; 254 T; 0 U; 2 Other;

Query Match 0.9%; Score 54.8; DB 7; Length 2060;
Best Local Similarity 46.5%; Pred. No. 0.00075;
Matches 213; Conservative 0; Mismatches 242; Indels 3; Gaps 1;

```
Qy       3348 CGGGAGTTTGGAACGCTCCACCTCGGGGAGGACCAGGATGAGGCCCAGGTCAACGGCCAC 3407
             || | | |||| ||||| |||| || ||| | |  || | | | || | |
Db        942 CGCGCGCTTGGCGCGCTCGCACTCGCGGCGGAGCTTGCCCAGCGCGCGCGCGTCGCCGGC 883

Qy       3408 GGTCTTGGTGTTGGGCATCACCCTTTTCCCCTGAGCCCAGAGGACCAGCACGTCCAGGGT 3467
             | | | |     | || ||   |  |  |||   |  || || |||  | |    ||
```

```
Db        882 GATGTCGCGCCCGTGCTTCCGGCGGATGACCTTGACGAAGTGGTCCATGAGGCGTTGGTC 823

Qy       3468 GGGG---CGGATTCCGTGGTCCTGGTAGGCCTTGACCCAGTTGAAGGAGAGCACGCCGTT 3524

Db        822 GAAGTCCTCGCCGCCGAGGTGGGTGTCGCCGTTGGTGGCAAGGACCTCGAACACGCCGTT 763

Qy       3525 GGCCATGTCTAGGGCGAGGTTCCTCCAGGGGTAGATGTAGTCGTCCAGGGTGAGCCTGGC 3584

Db        762 GTCGATGGCGAGGATGCTGACGTCGAACGTGCCGCCGCCGAGGTCGAAGACGAGGACGTT 703

Qy       3585 TTTCCCTGGCCTCAGCCGGACGGCCCAAAGGGTGCCGATGGCCCGGCGCTCCCCGTTGAC 3644

Db        702 CTTCTCCGCGCCCTTCTTGTCGATGCCGTAGGCGATGGCGGCGGCGGTCGGCTCGTTGAT 643

Qy       3645 GGTTTGGTGAAGGACGTCGGTGGCCACCAGGCCCTTTTTCTCAAGGACCTTCTTCCAGGC 3704

Db        642 GATGCGGTCGACGGTGAGCCCGGCGATGACGCCGGCGTCCTTGGTGGCCTGCCGCTGCGC 583

Qy       3705 GTGGACGGTCTGCCGGGTGACCCCCAGGTTGAGGGCCAACATCTCCAGGGGGACCATGAA 3764

Db        582 GTCGTTGAAGTAGGCCGGGACGGTGACGACGGCGCGCGTGACCTTCTCGCCGAGGTAGGC 523

Qy       3765 GACGACCGTCCCCACCTTCCTCCAGAGCTCCCGGTTGC 3802

Db        522 CTCGGCCGTCTCCTTCATCCGGGTGAGCACCATGGCGC 485
```

Search completed: March 17, 2004, 21:48:03
Job time : 1433.28 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: March 17, 2004, 21:15:22 ; Search time 273.095 Seconds
(without alignments)
11885.630 Million cell updates/sec

Title: US-09-664-186-6
Perfect score: 5849
Sequence: 1 tctagaaggtcagggtggac.....tgtagaggaccagaagaccc 5849

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	5849	100.0	5849	3	US-09-134-246-6	Sequence 6, Appli
c	2	1242	21.2	1242	2	US-08-960-756-1	Sequence 1, Appli
c	3	1026	17.5	1026	3	US-09-134-246-4	Sequence 4, Appli
	4	999	17.1	999	2	US-08-960-756-3	Sequence 3, Appli
c	5	180	3.1	180	3	US-09-134-246-5	Sequence 5, Appli
c	6	77	1.3	7218	1	US-08-232-463-14	Sequence 14, Appl
	7	56.4	1.0	7218	1	US-08-232-463-14	Sequence 14, Appl
c	8	51.8	0.9	1551	4	US-09-252-991A-5309	Sequence 5309, Ap
c	9	51.8	0.9	1563	4	US-09-252-991A-15934	Sequence 15934, A
	10	51.8	0.9	1953	4	US-09-252-991A-15763	Sequence 15763, A
c	11	51.8	0.9	3351	4	US-09-252-991A-15871	Sequence 15871, A
c	12	51.8	0.9	4158	4	US-09-252-991A-5348	Sequence 5348, Ap
	13	51.8	0.9	4953	4	US-09-252-991A-5227	Sequence 5227, Ap
	14	49.6	0.8	4403765	3	US-09-103-840A-2	Sequence 2, Appli
	15	49.6	0.8	4411529	3	US-09-103-840A-1	Sequence 1, Appli
c	16	46.8	0.8	2460	4	US-09-252-991A-2998	Sequence 2998, Ap
	17	46.8	0.8	2856	4	US-09-252-991A-2869	Sequence 2869, Ap
c	18	46.8	0.8	3387	4	US-09-252-991A-3101	Sequence 3101, Ap
	19	46.2	0.8	1473	4	US-09-152-060-43	Sequence 43, Appl
c	20	46	0.8	1572	4	US-09-252-991A-8725	Sequence 8725, Ap
c	21	46	0.8	1929	4	US-09-252-991A-8827	Sequence 8827, Ap
	22	46	0.8	2079	4	US-09-252-991A-9222	Sequence 9222, Ap
c	23	45.8	0.8	960	4	US-09-252-991A-6990	Sequence 6990, Ap
	24	45.8	0.8	1275	4	US-09-252-991A-7059	Sequence 7059, Ap
c	25	45.8	0.8	1926	4	US-09-249-585A-4	Sequence 4, Appli
c	26	45.8	0.8	1931	2	US-09-130-114-2	Sequence 2, Appli
c	27	45.8	0.8	2187	4	US-09-252-991A-7042	Sequence 7042, Ap
c	28	45.6	0.8	984	4	US-09-252-991A-10162	Sequence 10162, A
c	29	45.6	0.8	1491	4	US-09-252-991A-10097	Sequence 10097, A
c	30	45.6	0.8	2668	2	US-08-461-775-11	Sequence 11, Appl
c	31	45.6	0.8	2668	3	US-09-031-606-11	Sequence 11, Appl
	32	45.4	0.8	1758	4	US-09-252-991A-490	Sequence 490, App
c	33	45.4	0.8	1794	4	US-09-252-991A-568	Sequence 568, App
	34	45.4	0.8	3150	4	US-09-252-991A-11487	Sequence 11487, A
c	35	45.4	0.8	3372	4	US-09-252-991A-11405	Sequence 11405, A
	36	45.2	0.8	1215	4	US-09-252-991A-6208	Sequence 6208, Ap
c	37	45.2	0.8	1338	4	US-09-252-991A-6341	Sequence 6341, Ap
c	38	45.2	0.8	2034	4	US-09-252-991A-11958	Sequence 11958, A
c	39	45.2	0.8	2493	4	US-09-252-991A-11987	Sequence 11987, A
	40	45.2	0.8	2991	4	US-09-252-991A-12025	Sequence 12025, A
c	41	45	0.8	4767	4	US-09-231-899-76	Sequence 76, Appl
c	42	44.8	0.8	837	4	US-09-252-991A-4221	Sequence 4221, Ap
	43	44.8	0.8	1080	4	US-09-252-991A-4664	Sequence 4664, Ap
c	44	44.8	0.8	1440	4	US-09-252-991A-4726	Sequence 4726, Ap
c	45	44.8	0.8	2458	4	US-09-513-783A-175	Sequence 175, App

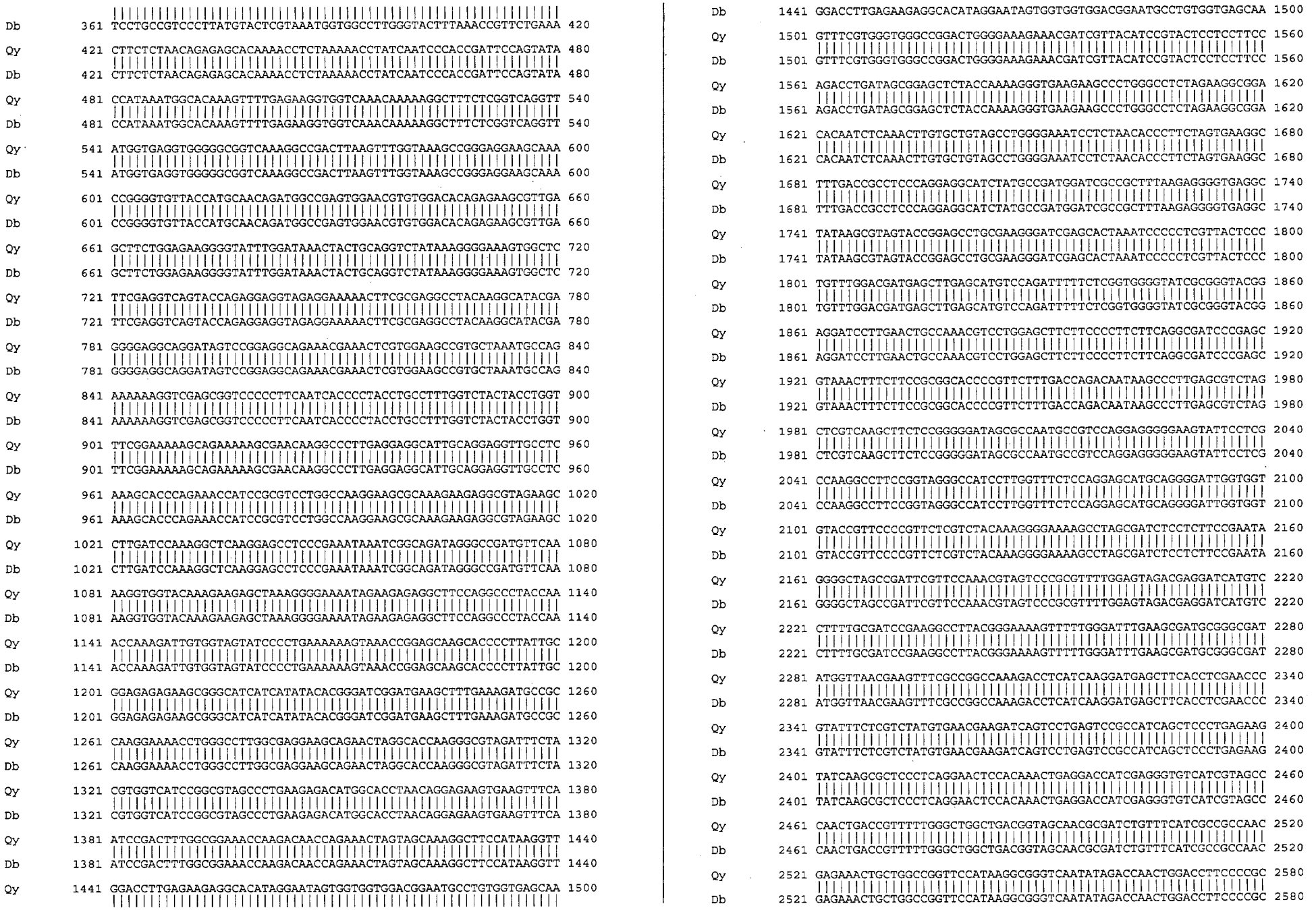
ALIGNMENTS

RESULT 1
US-09-134-246-6
; Sequence 6, Application US/09134246B
; Patent No. 6207377
; GENERAL INFORMATION:
; APPLICANT: Wayne, Jay
; APPLICANT: Xu, Shuang-yong
; TITLE OF INVENTION: Method For Construction Of Thermus-E. coli Shuttle
; TITLE OF INVENTION: Vectors And Identification of Two Thermus Plasmid
; TITLE OF INVENTION: Replication Origins
; FILE REFERENCE: Thermus Shuttle Vector
; CURRENT APPLICATION NUMBER: US/09/134,246B
; CURRENT FILING DATE: 1998-08-14
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 5849
; TYPE: DNA
; ORGANISM: Thermus sp.
US-09-134-246-6

Query Match 100.0%; Score 5849; DB 3; Length 5849;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 5849; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 TCTAGAAGGTCAGGGTGGACAAGGAAAACACCATAGCCCCTGCCAAGAAGATGGACGAGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TCTAGAAGGTCAGGGTGGACAAGGAAAACACCATAGCCCCTGCCAAGAAGATGGACGAGT 60

Qy         61 TGGTGTCCGGAAAAGTGGCCATCCGGGGCGCTCTTGACAACTATTTTCCAGCGGTGGCCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGGTGTCCGGAAAAGTGGCCATCCGGGGCGCTCTTGACAACTATTTTCCAGCGGTGGCCA 120

Qy        121 CCGGCATTGGCCACGAGGTACGAGCTTGTGGAGTAGACGGCCACAAAGGGGTCGTCCTCA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGCATTGGCCACGAGGTACGAGCTTGTGGAGTAGACGGCCACAAAGGGGTCGTCCTCA 180

Qy        181 AACTTCTTTTCTAGTGCCGCTTGGACGAAGGGGAGGAAGAGGAAAGGCTTCATGGCCTCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AACTTCTTTTCTAGTGCCGCTTGGACGAAGGGGAGGAAGAGGAAAGGCTTCATGGCCTCA 240

Qy        241 CCTCCTTCCCCTCCTCCTTGGCGGCCTTAGCGGCGTAAAACTCTGAGACGGCCTGAAGTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCTCCTTCCCCTCCTCCTTGGCGGCCTTAGCGGCGTAAAACTCTGAGACGGCCTGAAGTT 300

Qy        301 TAGGGATTTCGCTTTCGGGGATAAGAATCCGGCGGCTCAGGGGATGCCGGATGGCCCTTA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TAGGGATTTCGCTTTCGGGGATAAGAATCCGGCGGCTCAGGGGATGCCGGATGGCCCTTA 360

Qy        361 TCCTGCCGTCCCTTATGTACTCGTAAATGGTGGCCTTGGGTACTTTAAACCGTTCTGAAA 420
```

c	28	45.6	0.8	984	4	US-09-252-991A-10162	Sequence 10162, A
c	29	45.6	0.8	1491	4	US-09-252-991A-10097	Sequence 10097, A
c	30	45.6	0.8	2668	2	US-08-461-775-11	Sequence 11, Appl
c	31	45.6	0.8	2668	3	US-09-031-606-11	Sequence 11, Appl
	32	45.4	0.8	1758	4	US-09-252-991A-490	Sequence 490, App
c	33	45.4	0.8	1794	4	US-09-252-991A-568	Sequence 568, App
	34	45.4	0.8	3150	4	US-09-252-991A-11487	Sequence 11487, A
c	35	45.4	0.8	3372	4	US-09-252-991A-11405	Sequence 11405, A
	36	45.2	0.8	1215	4	US-09-252-991A-6208	Sequence 6208, Ap
c	37	45.2	0.8	1338	4	US-09-252-991A-6341	Sequence 6341, Ap
c	38	45.2	0.8	2034	4	US-09-252-991A-11958	Sequence 11958, A
c	39	45.2	0.8	2493	4	US-09-252-991A-11987	Sequence 11987, A
	40	45.2	0.8	2991	4	US-09-252-991A-12025	Sequence 12025, A
c	41	45	0.8	4767	4	US-09-231-899-76	Sequence 76, Appl
c	42	44.8	0.8	837	4	US-09-252-991A-4221	Sequence 4221, Ap
	43	44.8	0.8	1080	4	US-09-252-991A-4664	Sequence 4664, Ap
c	44	44.8	0.8	1440	4	US-09-252-991A-4726	Sequence 4726, Ap
c	45	44.8	0.8	2458	4	US-09-513-783A-175	Sequence 175, App


Db 361 TCCTGCCGTCCCTTATGTACTCGTAAATGGTGGCCTTGGGTACTTTAAACCGTTCTGAAA 420
Qy 421 CTTCTCTAACAGAGAGCACAAAACCTCTAAAAACCTATCAATCCCACCGATTCCAGTATA 480
Db 421 CTTCTCTAACAGAGAGCACAAAACCTCTAAAAACCTATCAATCCCACCGATTCCAGTATA 480
Qy 481 CCATAAATGGCACAAAGTTTTGAGAAGGTGGTCAAACAAAAAGGCTTTCTCGGTCAGGTT 540
Db 481 CCATAAATGGCACAAAGTTTTGAGAAGGTGGTCAAACAAAAAGGCTTTCTCGGTCAGGTT 540
Qy 541 ATGGTGAGGTGGGGGCGGTCAAAGGCCGACTTAAGTTTGGTAAAGCCGGGAGGAAGCAAA 600
Db 541 ATGGTGAGGTGGGGGCGGTCAAAGGCCGACTTAAGTTTGGTAAAGCCGGGAGGAAGCAAA 600
Qy 601 CCGGGGTGTTACCATGCAACAGATGGCCGAGTGGAACGTGTGGACACAGAGAAGCGTTGA 660
Db 601 CCGGGGTGTTACCATGCAACAGATGGCCGAGTGGAACGTGTGGACACAGAGAAGCGTTGA 660
Qy 661 GCTTCTGGAGAAGGGGTATTTGGATAAACTACTGCAGGTCTATAAAGGGGAAAGTGGCTC 720
Db 661 GCTTCTGGAGAAGGGGTATTTGGATAAACTACTGCAGGTCTATAAAGGGGAAAGTGGCTC 720
Qy 721 TTCGAGGTCAGTACCAGAGGAGGTAGAGGAAAAACTTCGCGAGGCCTACAAGGCATACGA 780
Db 721 TTCGAGGTCAGTACCAGAGGAGGTAGAGGAAAAACTTCGCGAGGCCTACAAGGCATACGA 780
Qy 781 GGGGAGGCAGGATAGTCCGGAGGCAGAAACGAAACTCGTGGAAGCCGTGCTAAATGCCAG 840
Db 781 GGGGAGGCAGGATAGTCCGGAGGCAGAAACGAAACTCGTGGAAGCCGTGCTAAATGCCAG 840
Qy 841 AAAAAAGGTCGAGCGGTCCCCCTTCAATCACCCCTACCTGCCTTTGGTCTACTACCTGGT 900
Db 841 AAAAAAGGTCGAGCGGTCCCCCTTCAATCACCCCTACCTGCCTTTGGTCTACTACCTGGT 900
Qy 901 TTCGGAAAAAGCAGAAAAAGCGAACAAGGCCCTTGAGGAGGCATTGCAGGAGGTTGCCTC 960
Db 901 TTCGGAAAAAGCAGAAAAAGCGAACAAGGCCCTTGAGGAGGCATTGCAGGAGGTTGCCTC 960
Qy 961 AAAGCACCCAGAAACCATCCGCGTCCTGGCCAAGGAAGCGCAAAGAAGAGGCGTAGAAGC 1020
Db 961 AAAGCACCCAGAAACCATCCGCGTCCTGGCCAAGGAAGCGCAAAGAAGAGGCGTAGAAGC 1020
Qy 1021 CTTGATCCAAAGGCTCAAGGAGCCTCCCGAAATAAATCGGCAGATAGGGCCGATGTTCAA 1080
Db 1021 CTTGATCCAAAGGCTCAAGGAGCCTCCCGAAATAAATCGGCAGATAGGGCCGATGTTCAA 1080
Qy 1081 AAGGTGGTACAAAGAAGAGCTAAAGGGGAAAATAGAAGAGAGGCTTCCAGGCCCTACCAA 1140
Db 1081 AAGGTGGTACAAAGAAGAGCTAAAGGGGAAAATAGAAGAGAGGCTTCCAGGCCCTACCAA 1140
Qy 1141 ACCAAAGATTGTGGTAGTATCCCCTGAAAAAAGTAAACCGGAGCAAGCACCCCTTATTGC 1200
Db 1141 ACCAAAGATTGTGGTAGTATCCCCTGAAAAAAGTAAACCGGAGCAAGCACCCCTTATTGC 1200
Qy 1201 GGAGAGAGAAGCGGGCATCATCATATACACGGGATCGGATGAAGCTTTGAAAGATGCCGC 1260
Db 1201 GGAGAGAGAAGCGGGCATCATCATATACACGGGATCGGATGAAGCTTTGAAAGATGCCGC 1260
Qy 1261 CAAGGAAAACCTGGGCCTTGGCGAGGAAGCAGAACTAGGCACCAAGGGCGTAGATTTCTA 1320
Db 1261 CAAGGAAAACCTGGGCCTTGGCGAGGAAGCAGAACTAGGCACCAAGGGCGTAGATTTCTA 1320
Qy 1321 CGTGGTCATCCGGCGTAGCCCTGAAGAGACATGGCACCTAACAGGAGAAGTGAAGTTTCA 1380
Db 1321 CGTGGTCATCCGGCGTAGCCCTGAAGAGACATGGCACCTAACAGGAGAAGTGAAGTTTCA 1380
Qy 1381 ATCCGACTTTGGCGGAAACCAAGACAACCAGAAACTAGTAGCAAAGGCTTCCATAAGGTT 1440
Db 1381 ATCCGACTTTGGCGGAAACCAAGACAACCAGAAACTAGTAGCAAAGGCTTCCATAAGGTT 1440
Qy 1441 GGACCTTGAGAAGAGGCACATAGGAATAGTGGTGGTGGACGGAATGCCTGTGGTGAGCAA 1500
Db 1441 GGACCTTGAGAAGAGGCACATAGGAATAGTGGTGGTGGACGGAATGCCTGTGGTGAGCAA 1500
Qy 1501 GTTTCGTGGGTGGGCCGGACTGGGGAAAGAAACGATCGTTACATCCGTACTCCTCCTTCC 1560
Db 1501 GTTTCGTGGGTGGGCCGGACTGGGGAAAGAAACGATCGTTACATCCGTACTCCTCCTTCC 1560
Qy 1561 AGACCTGATAGCGGAGCTCTACCAAAAGGGTGAAGAAGCCCTGGGCCTCTAGAAGGCGGA 1620
Db 1561 AGACCTGATAGCGGAGCTCTACCAAAAGGGTGAAGAAGCCCTGGGCCTCTAGAAGGCGGA 1620
Qy 1621 CACAATCTCAAACTTGTGCTGTAGCCTGGGGAAATCCTCTAACACCCTTCTAGTGAAGGC 1680
Db 1621 CACAATCTCAAACTTGTGCTGTAGCCTGGGGAAATCCTCTAACACCCTTCTAGTGAAGGC 1680
Qy 1681 TTTGACCGCCTCCCAGGAGGCATCTATGCCGATGGATCGCCGCTTTAAGAGGGGTGAGGC 1740
Db 1681 TTTGACCGCCTCCCAGGAGGCATCTATGCCGATGGATCGCCGCTTTAAGAGGGGTGAGGC 1740
Qy 1741 TATAAGCGTAGTACCGGAGCCTGCGAAGGGATCGAGCACTAAATCCCCCTCGTTACTCCC 1800
Db 1741 TATAAGCGTAGTACCGGAGCCTGCGAAGGGATCGAGCACTAAATCCCCCTCGTTACTCCC 1800
Qy 1801 TGTTTGGACGATGAGCTTGAGCATGTCCAGATTTTTCTCGGTGGGGTATCGCGGGTACGG 1860
Db 1801 TGTTTGGACGATGAGCTTGAGCATGTCCAGATTTTTCTCGGTGGGGTATCGCGGGTACGG 1860
Qy 1861 AGGATCCTTGAACTGCCAAACGTCCTGGAGCTTCTTCCCCTTCTTCAGGCGATCCCGAGC 1920
Db 1861 AGGATCCTTGAACTGCCAAACGTCCTGGAGCTTCTTCCCCTTCTTCAGGCGATCCCGAGC 1920
Qy 1921 GTAAACTTTCTTCCGCGGCACCCCGTTCTTTGACCAGACAATAAGCCCTTGAGCGTCTAG 1980
Db 1921 GTAAACTTTCTTCCGCGGCACCCCGTTCTTTGACCAGACAATAAGCCCTTGAGCGTCTAG 1980
Qy 1981 CTCGTCAAGCTTCTCCGGGGGATAGCGCCAATGCCGTCCAGGAGGGGGAAGTATTCCTCG 2040
Db 1981 CTCGTCAAGCTTCTCCGGGGGATAGCGCCAATGCCGTCCAGGAGGGGGAAGTATTCCTCG 2040
Qy 2041 CCAAGGCCTTCCGGTAGGGCCATCCTTGGTTTCTCCAGGAGCATGCAGGGGATTGGTGGT 2100
Db 2041 CCAAGGCCTTCCGGTAGGGCCATCCTTGGTTTCTCCAGGAGCATGCAGGGGATTGGTGGT 2100
Qy 2101 GTACCGTTCCCCGTTCTCGTCTACAAAGGGGAAAAGCCTAGCGATCTCCTCTTCCGAATA 2160
Db 2101 GTACCGTTCCCCGTTCTCGTCTACAAAGGGGAAAAGCCTAGCGATCTCCTCTTCCGAATA 2160
Qy 2161 GGGGCTAGCCGATTCGTTCCAAACGTAGTCCCGCGTTTTGGAGTAGACGAGGATCATGTC 2220
Db 2161 GGGGCTAGCCGATTCGTTCCAAACGTAGTCCCGCGTTTTGGAGTAGACGAGGATCATGTC 2220
Qy 2221 CTTTTGCGATCCGAAGGCCTTACGGGAAAAGTTTTTGGGATTTGAAGCGATGCGGGCGAT 2280
Db 2221 CTTTTGCGATCCGAAGGCCTTACGGGAAAAGTTTTTGGGATTTGAAGCGATGCGGGCGAT 2280
Qy 2281 ATGGTTAACGAAGTTTCGCCGGCCAAAGACCTCATCAAGGATGAGCTTCACCTCGAACCC 2340
Db 2281 ATGGTTAACGAAGTTTCGCCGGCCAAAGACCTCATCAAGGATGAGCTTCACCTCGAACCC 2340
Qy 2341 GTATTTCTCGTCTATGTGAACGAAGATCAGTCCTGAGTCCGCCATCAGCTCCCTGAGAAG 2400
Db 2341 GTATTTCTCGTCTATGTGAACGAAGATCAGTCCTGAGTCCGCCATCAGCTCCCTGAGAAG 2400
Qy 2401 TATCAAGCGCTCCCTCAGGAACTCCACAAACTGAGGACCATCGAGGGTGTCATCGTAGCC 2460
Db 2401 TATCAAGCGCTCCCTCAGGAACTCCACAAACTGAGGACCATCGAGGGTGTCATCGTAGCC 2460
Qy 2461 CAACTGACCGTTTTTGGGCTGGCTGACGGTAGCAACGCGATCTGTTTCATCGCCGCCAAC 2520
Db 2461 CAACTGACCGTTTTTGGGCTGGCTGACGGTAGCAACGCGATCTGTTTCATCGCCGCCAAC 2520
Qy 2521 GAGAAACTGCTGGCCGGTTCCATAAGGCGGGTCAATATAGACCAACTGGACCTTCCCCGC 2580
Db 2521 GAGAAACTGCTGGCCGGTTCCATAAGGCGGGTCAATATAGACCAACTGGACCTTCCCCGC 2580

```
Qy       2581 ATACCCACCAGGCTCCCGGAGCATCCACCGGAGAACCTGACCGTTTTCCCCCAAAAAGTA 2640
Db       2581 ATACCCACCAGGCTCCCGGAGCATCCACCGGAGAACCTGACCGTTTTCCCCCAAAAAGTA 2640

Qy       2641 GGTGCCAATAGGATCAATCTCAAAAAGGGGGGCATTTCCCCCTAGGAAGAGGAGGGTTTC 2700
Db       2641 GGTGCCAATAGGATCAATCTCAAAAAGGGGGGCATTTCCCCCTAGGAAGAGGAGGGTTTC 2700

Qy       2701 TTTTCGCAAAACAAGTTGTGGGGTGGGCTGATCAAGAATCTCCTTCTCATCGCGTTTTCC 2760
Db       2701 TTTTCGCAAAACAAGTTGTGGGGTGGGCTGATCAAGAATCTCCTTCTCATCGCGTTTTCC 2760

Qy       2761 GGGGTAGACCAACCTAAAGGGCGAAGGTTCCGAGGTTTTCGAGGCTTTCAAGGGGGCTTT 2820
Db       2761 GGGGTAGACCAACCTAAAGGGCGAAGGTTCCGAGGTTTTCGAGGCTTTCAAGGGGGCTTT 2820

Qy       2821 TCGGGTCAAACCAGGGTAGCTACGGCTCATTCTTCCCTCCCCACAGCGCTCTTAAGCAGG 2880
Db       2821 TCGGGTCAAACCAGGGTAGCTACGGCTCATTCTTCCCTCCCCACAGCGCTCTTAAGCAGG 2880

Qy       2881 ACCTCATCACCCACAACCCTCACGCACTCCAACCAAGGAATCCGCCAAAGGCGGCCTACC 2940
Db       2881 ACCTCATCACCCACAACCCTCACGCACTCCAACCAAGGAATCCGCCAAAGGCGGCCTACC 2940

Qy       2941 TTTTGAGCCCGTATCTTCCCCTGACGTATAGACCTTCGGATCGTCTCAGGGTGCACCCGA 3000
Db       2941 TTTTGAGCCCGTATCTTCCCCTGACGTATAGACCTTCGGATCGTCTCAGGGTGCACCCGA 3000

Qy       3001 AGGATGTCTGCAAGCTCCTCGGGGGTCAGGTACACGGGCTTCATCCTCATGACACAACCT 3060
Db       3001 AGGATGTCTGCAAGCTCCTCGGGGGTCAGGTACACGGGCTTCATCCTCATGACACAACCT 3060

Qy       3061 TACCCCACAGAGGACAACACATGCAACTATGGGCAAAGTAGACAACGAGACCAAAAGCTT 3120
Db       3061 TACCCCACAGAGGACAACACATGCAACTATGGGCAAAGTAGACAACGAGACCAAAAGCTT 3120

Qy       3121 GGGCCACTCTCTCAGGAGGCCTCCTTGAGGGTCTTCACTAGGTACGCTCCCGGTCGTGTC 3180
Db       3121 GGGCCACTCTCTCAGGAGGCCTCCTTGAGGGTCTTCACTAGGTACGCTCCCGGTCGTGTC 3180

Qy       3181 AGATGGCCATCCGTGTAATCTCGGATAACCCGCATTAGGACGGCAAATAGATATTGCGCG 3240
Db       3181 AGATGGCCATCCGTGTAATCTCGGATAACCCGCATTAGGACGGCAAATAGATATTGCGCG 3240

Qy       3241 GGGAGTTCACCCCTGGCCACAGCCCACAGCAAGCCTGCATAGAAACGTCTTGAACGACGG 3300
Db       3241 GGGAGTTCACCCCTGGCCACAGCCCACAGCAAGCCTGCATAGAAACGTCTTGAACGACGG 3300

Qy       3301 TCATCTAGGAGATCGGCAATGTACGTAGCAATGAGGGTGATAAGGGCCGGGAGTTTGGAA 3360
Db       3301 TCATCTAGGAGATCGGCAATGTACGTAGCAATGAGGGTGATAAGGGCCGGGAGTTTGGAA 3360

Qy       3361 CGCTCCACCTCGGGGAGGACCAGGATGAGGCCCAGGTCAACGGCCACGGTCTTGGTGTTG 3420
Db       3361 CGCTCCACCTCGGGGAGGACCAGGATGAGGCCCAGGTCAACGGCCACGGTCTTGGTGTTG 3420

Qy       3421 GGCATCACCCTTTTCCCCTGAGCCCAGAGGACCAGCACGTCCAGGGTGGGGCGGATTCCG 3480
Db       3421 GGCATCACCCTTTTCCCCTGAGCCCAGAGGACCAGCACGTCCAGGGTGGGGCGGATTCCG 3480

Qy       3481 TGGTCCTGGTAGGCCTTGACCCAGTTGAAGGAGAGCACGCCGTTGGCCATGTCTAGGGCG 3540
Db       3481 TGGTCCTGGTAGGCCTTGACCCAGTTGAAGGAGAGCACGCCGTTGGCCATGTCTAGGGCG 3540

Qy       3541 AGGTTCCTCCAGGGGTAGATGTAGTCGTCCAGGGTGAGCCTGGCTTTCCCTGGCCTCAGC 3600
Db       3541 AGGTTCCTCCAGGGGTAGATGTAGTCGTCCAGGGTGAGCCTGGCTTTCCCTGGCCTCAGC 3600

Qy       3601 CGGACGGCCCAAAGGGTGCCGATGGCCCGGCGCTCCCCGTTGACGGTTTGGTGAAGGACG 3660
Db       3601 CGGACGGCCCAAAGGGTGCCGATGGCCCGGCGCTCCCCGTTGACGGTTTGGTGAAGGACG 3660

Qy       3661 TCGGTGGCCACCAGGCCCTTTTTCTCAAGGACCTTCTTCCAGGCGTGGACGGTCTGCCGG 3720
Db       3661 TCGGTGGCCACCAGGCCCTTTTTCTCAAGGACCTTCTTCCAGGCGTGGACGGTCTGCCGG 3720

Qy       3721 GTGACCCCCAGGTTGAGGGCCAACATCTCCAGGGGGACCATGAAGACGACCGTCCCCACC 3780
Db       3721 GTGACCCCCAGGTTGAGGGCCAACATCTCCAGGGGGACCATGAAGACGACCGTCCCCACC 3780

Qy       3781 TTCCTCCAGAGCTCCCGGTTGCCGTAGGGGATGGTGGAGCGGGCAATCTCCTGGAGGAGT 3840
Db       3781 TTCCTCCAGAGCTCCCGGTTGCCGTAGGGGATGGTGGAGCGGGCAATCTCCTGGAGGAGT 3840

Qy       3841 TCCAGAAGCTTCTGAGGCCCGTCCTGGACGGCTTGACATACATTCCCAACGGGGGGTTCA 3900
Db       3841 TCCAGAAGCTTCTGAGGCCCGTCCTGGACGGCTTGACATACATTCCCAACGGGGGGTTCA 3900

Qy       3901 GCTGGTGGTATAGGGGGGGCCCCAACAGAGGAAAGGAAAGCCTCATGGTCTTTCTCGGGT 3960
Db       3901 GCTGGTGGTATAGGGGGGGCCCCAACAGAGGAAAGGAAAGCCTCATGGTCTTTCTCGGGT 3960

Qy       3961 AAAGCAATCCTAAGTGCCTCTTTTTGGTATGTAAAGCCCTTCGCGAGGCGATTTTCGGCA 4020
Db       3961 AAAGCAATCCTAAGTGCCTCTTTTTGGTATGTAAAGCCCTTCGCGAGGCGATTTTCGGCA 4020

Qy       4021 CCTCCATCTGGAGGGGGGTCGGTGGCCAAGAAGAAGTCCTCTGACCCCCTATCTGACCCC 4080
Db       4021 CCTCCATCTGGAGGGGGGTCGGTGGCCAAGAAGAAGTCCTCTGACCCCCTATCTGACCCC 4080

Qy       4081 CTAGTGGCATCGGTGTTGTCGTGGGTTTCCTCTAAAGCCTCGTAAAGCTCTTCAAAGAAG 4140
Db       4081 CTAGTGGCATCGGTGTTGTCGTGGGTTTCCTCTAAAGCCTCGTAAAGCTCTTCAAAGAAG 4140

Qy       4141 GTTTTTTCGTTCTTCACCCTCGGACCTCCTTGTCATCTGGAGCCCGAGGCGTTACCCTAG 4200
Db       4141 GTTTTTTCGTTCTTCACCCTCGGACCTCCTTGTCATCTGGAGCCCGAGGCGTTACCCTAG 4200

Qy       4201 GTCTTGGGGGTGATCCGGGGCAACCGCCTCGGTTTCGCCTTTTTATGGGTCCAAAATAAC 4260
Db       4201 GTCTTGGGGGTGATCCGGGGCAACCGCCTCGGTTTCGCCTTTTTATGGGTCCAAAATAAC 4260

Qy       4261 CGTCAGCCCAGCGGCTGGCAATCCCCCCTCCTAAAAGGCCGTTATAGGCCCTGCTAGGAG 4320
Db       4261 CGTCAGCCCAGCGGCTGGCAATCCCCCCTCCTAAAAGGCCGTTATAGGCCCTGCTAGGAG 4320

Qy       4321 GGGGGTAGTACTTTCCTACCCCCCTAGGCTTGGAGAGGCCTTAGGAGGTCTCCTAGGGCC 4380
Db       4321 GGGGGTAGTACTTTCCTACCCCCCTAGGCTTGGAGAGGCCTTAGGAGGTCTCCTAGGGCC 4380

Qy       4381 TCGTGGGGGTGTAGGGGTAACCTCATGGCCAGGCCGGCCGGCTCGGGACTCTGGAGGAGG 4440
Db       4381 TCGTGGGGGTGTAGGGGTAACCTCATGGCCAGGCCGGCCGGCTCGGGACTCTGGAGGAGG 4440

Qy       4441 CCTCCATAGCCTACTCGTGGTGGAGGTTTGTGAAGGGGTTCACTAATGCATACGGCTAGC 4500
Db       4441 CCTCCATAGCCTACTCGTGGTGGAGGTTTGTGAAGGGGTTCACTAATGCATACGGCTAGC 4500

Qy       4501 CTCGGGATCACGGCCAAATGGTATGCAGGTTTTGGTATAAAACCCTCAGGTTTGAGGCTA 4560
Db       4501 CTCGGGATCACGGCCAAATGGTATGCAGGTTTTGGTATAAAACCCTCAGGTTTGAGGCTA 4560

Qy       4561 GTTTATGTCGGTTTTATGCACCTTTGACTCGGATCACGGGCATAAACACCAGTTTCCTGC 4620
Db       4561 GTTTATGTCGGTTTTATGCACCTTTGACTCGGATCACGGGCATAAACACCAGTTTCCTGC 4620

Qy       4621 ACGAAAGAAAACTTTCGCGATCTAAGAGGGGGAAAGAGGTGTAGAGGGACGGCCTTCATG 4680
Db       4621 ACGAAAGAAAACTTTCGCGATCTAAGAGGGGGAAAGAGGTGTAGAGGGACGGCCTTCATG 4680

Qy       4681 AAAGTTGGCCTCTTAGGAGGCCGTTGTAGAGGGCCGTCTCGGGTTCAAATCCTTTCCCTC 4740
Db       4681 AAAGTTGGCCTCTTAGGAGGCCGTTGTAGAGGGCCGTCTCGGGTTCAAATCCTTTCCCTC 4740

Qy       4741 TCTCTCCAGGTTTCCGAGGTTCGAGGTCTTGGTCCAGGTCTTGTACCAAGTTTTTGACCA 4800
```

```
Db       4741 TCTCTCCAGGTTTCCGAGGTTCGAGGTCTTGGTCCAGGTCTTGTACCAAGTTTTTGACCA 4800

Qy       4801 AAGTCTATTCTCGGAATATAGGGGTATCTTGTCTATCTTCCCTACGGGATATCTCTGTCT 4860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4801 AAGTCTATTCTCGGAATATAGGGGTATCTTGTCTATCTTCCCTACGGGATATCTCTGTCT 4860

Qy       4861 GTGTGAACTTGATCCCATCCCAATACATATCTCAATCTCCTAATCTCCTCTTCTCTCCAG 4920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4861 GTGTGAACTTGATCCCATCCCAATACATATCTCAATCTCCTAATCTCCTCTTCTCTCCAG 4920

Qy       4921 ATCCCTAATCTCTTCTTCTACCTCTTTCTCCTCCCAATTAAGAATGGAGAGGAAAAACCC 4980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4921 ATCCCTAATCTCTTCTTCTACCTCTTTCTCCTCCCAATTAAGAATGGAGAGGAAAAACCC 4980

Qy       4981 CGACCAGAACGAGCTTCTCGGGGTCAGTTTCGGTAATCTCGGGACAGGTTTTCATCGTCT 5040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4981 CGACCAGAACGAGCTTCTCGGGGTCAGTTTCGGTAATCTCGGGACAGGTTTTCATCGTCT 5040

Qy       5041 AGGACGAGGATTAGGGCATGAAAAATGGGCTTTGACAAAATCTTTCTAAAAAATACTCCC 5100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5041 AGGACGAGGATTAGGGCATGAAAAATGGGCTTTGACAAAATCTTTCTAAAAAATACTCCC 5100

Qy       5101 CGAGGTTGGGGAAGTGCCCTCGGGGAGAAGATTTTTGGCAGTTTAGATGTTATGCTCTAT 5160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5101 CGAGGTTGGGGAAGTGCCCTCGGGGAGAAGATTTTTGGCAGTTTAGATGTTATGCTCTAT 5160

Qy       5161 CACGGGCCGGAGGCCTCCACGATAAGTTGTCTTGGCCAAGTACCGGGCCAGGTCGGGGGT 5220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5161 CACGGGCCGGAGGCCTCCACGATAAGTTGTCTTGGCCAAGTACCGGGCCAGGTCGGGGGT 5220

Qy       5221 GCTCTTCAGCGTGGTGATGGTACTTTCACGGAAGTTCACAAGTCCTTTTAGAGGCTTCAG 5280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5221 GCTCTTCAGCGTGGTGATGGTACTTTCACGGAAGTTCACAAGTCCTTTTAGAGGCTTCAG 5280

Qy       5281 GTCGGGGATAGTGCTCAAGTACTCCCAAGCGTTCTCGGGCCCGTGGTCGGGGAGAAGGAC 5340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5281 GTCGGGGATAGTGCTCAAGTACTCCCAAGCGTTCTCGGGCCCGTGGTCGGGGAGAAGGAC 5340

Qy       5341 AAAGGGGTCGGGCAAAAGTTCATCTTTGTACTTAGGACGGATTACTTTAGCACCTGATAA 5400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5341 AAAGGGGTCGGGCAAAAGTTCATCTTTGTACTTAGGACGGATTACTTTAGCACCTGATAA 5400

Qy       5401 CTTCAGGGCCGTTAAGAAGGGCCTCACCTCGGAGACGGGTGGAAGGAGGACGTGGGCGTG 5460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5401 CTTCAGGGCCGTTAAGAAGGGCCTCACCTCGGAGACGGGTGGAAGGAGGACGTGGGCGTG 5460

Qy       5461 GAAGAAGACGAACCCCGATTTTTGGGAAGTCTCCCTCCAGTTTGATGATGAACGTTGGGA 5520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5461 GAAGAAGACGAACCCCGATTTTTGGGAAGTCTCCCTCCAGTTTGATGATGAACGTTGGGA 5520

Qy       5521 GGAAGCCGGCCAGGATGTCTTTCATCGCGCCTCGAACCTCGGACACATAAAAAACTTTCG 5580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5521 GGAAGCCGGCCAGGATGTCTTTCATCGCGCCTCGAACCTCGGACACATAAAAAACTTTCG 5580

Qy       5581 TGTTTGTCAGGGCAAGAGTGCTATGTATGAGGTAACCTTCGGGAGTACAAAGTGCCTCAA 5640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5581 TGTTTGTCAGGGCAAGAGTGCTATGTATGAGGTAACCTTCGGGAGTACAAAGTGCCTCAA 5640

Qy       5641 GCCGCCTTTCCCAACGCTCCAAAACTCTAGGGTCAGGTGGTTTAGGTTTTCTGAAAAACT 5700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5641 GCCGCCTTTCCCAACGCTCCAAAACTCTAGGGTCAGGTGGTTTAGGTTTTCTGAAAAACT 5700

Qy       5701 CTAGCTTTTCAGTGGTCATTCCTCACCCCTCTAGCACGTACTCTGGAAGGTAAACCTTTG 5760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5701 CTAGCTTTTCAGTGGTCATTCCTCACCCCTCTAGCACGTACTCTGGAAGGTAAACCTTTG 5760

Qy       5761 ACACAGCGGCCAAGTCTAGCGTCTCCCAGTCCAGTTGGTCTGGGACGCGTGAGAAGGGGA 5820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5761 ACACAGCGGCCAAGTCTAGCGTCTCCCAGTCCAGTTGGTCTGGGACGCGTGAGAAGGGGA 5820

Qy       5821 GGGGCTTGGTGTAGAGGACCAGAAGACCC 5849
             ||||||||||||||||||||||||||||
```

```
Db       5821 GGGGCTTGGTGTAGAGGACCAGAAGACCC 5849
```

```
RESULT 2
US-08-960-756-1/c
; Sequence 1, Application US/08960756
; Patent No. 5866422
;  GENERAL INFORMATION:
;    APPLICANT:  WAYNE, JAY
;    APPLICANT:  XU, SHUANG-YONG
;    TITLE OF INVENTION:  METHOD FOR CLONING AND
;    TITLE OF INVENTION:  PRODUCING THE Tsp45I RESTRICTION ENDONUCLEASE IN E. COLI
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  New England Biolabs, Inc.
;      STREET:  32 Tozer Road
;      CITY:  Beverly
;      STATE:  MA
;      COUNTRY:  US
;      ZIP:  01915
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/960,756
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Williams, Gregory D
;      REGISTRATION NUMBER:  30901
;      REFERENCE/DOCKET NUMBER:  NEB-128
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  978-927-5054
;      TELEFAX:  978-927-1705
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1242 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    FEATURE:
;      NAME/KEY:  Coding Sequence
;      LOCATION:  1...1239
;      OTHER INFORMATION:
US-08-960-756-1

  Query Match           21.2%;  Score 1242;  DB 2;  Length 1242;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1242;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1609 CTAGAAGGCGGACACAATCTCAAACTTGTGCTGTAGCCTGGGGAAATCCTCTAACACCCT 1668
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1242 CTAGAAGGCGGACACAATCTCAAACTTGTGCTGTAGCCTGGGGAAATCCTCTAACACCCT 1183

Qy       1669 TCTAGTGAAGGCTTTGACCGCCTCCCAGGAGGCATCTATGCCGATGGATCGCCGCTTTAA 1728
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1182 TCTAGTGAAGGCTTTGACCGCCTCCCAGGAGGCATCTATGCCGATGGATCGCCGCTTTAA 1123

Qy       1729 GAGGGGTGAGGCTATAAGCGTAGTACCGGAGCCTGCGAAGGGATCGAGCACTAAATCCCC 1788
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1122 GAGGGGTGAGGCTATAAGCGTAGTACCGGAGCCTGCGAAGGGATCGAGCACTAAATCCCC 1063

Qy       1789 CTCGTTACTCCCTGTTTGGACGATGAGCTTGAGCATGTCCAGATTTTTCTCGGTGGGGTA 1848
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1062 CTCGTTACTCCCTGTTTGGACGATGAGCTTGAGCATGTCCAGATTTTTCTCGGTGGGGTA 1003
```

```
Qy       1849 TCGCGGGTACGGAGGATCCTTGAACTGCCAAACGTCCTGGAGCTTCTTCCCCTTCTTCAG 1908
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1002 TCGCGGGTACGGAGGATCCTTGAACTGCCAAACGTCCTGGAGCTTCTTCCCCTTCTTCAG 943

Qy       1909 GCGATCCCGAGCGTAAACTTTCTTCCGCGGCACCCCGTTCTTTGACCAGACAATAAGCCC 1968
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        942 GCGATCCCGAGCGTAAACTTTCTTCCGCGGCACCCCGTTCTTTGACCAGACAATAAGCCC 883

Qy       1969 TTGAGCGTCTAGCTCGTCAAGCTTCTCCGGGGGATAGCGCCAATGCCGTCCAGGAGGGGG 2028
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        882 TTGAGCGTCTAGCTCGTCAAGCTTCTCCGGGGGATAGCGCCAATGCCGTCCAGGAGGGGG 823

Qy       2029 AAGTATTCCTCGCCAAGGCCTTCCGGTAGGGCCATCCTTGGTTTCTCCAGGAGCATGCAG 2088
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        822 AAGTATTCCTCGCCAAGGCCTTCCGGTAGGGCCATCCTTGGTTTCTCCAGGAGCATGCAG 763

Qy       2089 GGGATTGGTGGTGTACCGTTCCCCGTTCTCGTCTACAAAGGGGAAAAGCCTAGCGATCTC 2148
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        762 GGGATTGGTGGTGTACCGTTCCCCGTTCTCGTCTACAAAGGGGAAAAGCCTAGCGATCTC 703

Qy       2149 CTCTTCCGAATAGGGGCTAGCCGATTCGTTCCAAACGTAGTCCCGCGTTTTGGAGTAGAC 2208
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        702 CTCTTCCGAATAGGGGCTAGCCGATTCGTTCCAAACGTAGTCCCGCGTTTTGGAGTAGAC 643

Qy       2209 GAGGATCATGTCCTTTTGCGATCCGAAGGCCTTACGGGAAAAGTTTTTGGGATTTGAAGC 2268
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        642 GAGGATCATGTCCTTTTGCGATCCGAAGGCCTTACGGGAAAAGTTTTTGGGATTTGAAGC 583

Qy       2269 GATGCGGGCGATATGGTTAACGAAGTTTCGCCGGCCAAAGACCTCATCAAGGATGAGCTT 2328
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        582 GATGCGGGCGATATGGTTAACGAAGTTTCGCCGGCCAAAGACCTCATCAAGGATGAGCTT 523

Qy       2329 CACCTCGAACCCGTATTTCTCGTCTATGTGAACGAAGATCAGTCCTGAGTCCGCCATCAG 2388
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        522 CACCTCGAACCCGTATTTCTCGTCTATGTGAACGAAGATCAGTCCTGAGTCCGCCATCAG 463

Qy       2389 CTCCCTGAGAAGTATCAAGCGCTCCCTCAGGAACTCCACAAACTGAGGACCATCGAGGGT 2448
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        462 CTCCCTGAGAAGTATCAAGCGCTCCCTCAGGAACTCCACAAACTGAGGACCATCGAGGGT 403

Qy       2449 GTCATCGTAGCCCAACTGACCGTTTTTGGGCTGGCTGACGGTAGCAACGCGATCTGTTTC 2508
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        402 GTCATCGTAGCCCAACTGACCGTTTTTGGGCTGGCTGACGGTAGCAACGCGATCTGTTTC 343

Qy       2509 ATCGCCGCCAACGAGAAACTGCTGGCCGGTTCCATAAGGCGGGTCAATATAGACCAACTG 2568
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        342 ATCGCCGCCAACGAGAAACTGCTGGCCGGTTCCATAAGGCGGGTCAATATAGACCAACTG 283

Qy       2569 GACCTTCCCCGCATACCCACCAGGCTCCCGGAGCATCCACCGGAGAACCTGACCGTTTTC 2628
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        282 GACCTTCCCCGCATACCCACCAGGCTCCCGGAGCATCCACCGGAGAACCTGACCGTTTTC 223

Qy       2629 CCCCAAAAAGTAGGTGCCAATAGGATCAATCTCAAAAAGGGGGGCATTTCCCCCTAGGAA 2688
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        222 CCCCAAAAAGTAGGTGCCAATAGGATCAATCTCAAAAAGGGGGGCATTTCCCCCTAGGAA 163

Qy       2689 GAGGAGGGTTTCTTTTCGCAAAACAAGTTGTGGGGTGGGCTGATCAAGAATCTCCTTCTC 2748
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        162 GAGGAGGGTTTCTTTTCGCAAAACAAGTTGTGGGGTGGGCTGATCAAGAATCTCCTTCTC 103

Qy       2749 ATCGCGTTTTCCGGGGTAGACCAACCTAAAGGGCGAAGGTTCCGAGGTTTTCGAGGCTTT 2808
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        102 ATCGCGTTTTCCGGGGTAGACCAACCTAAAGGGCGAAGGTTCCGAGGTTTTCGAGGCTTT 43

Qy       2809 CAAGGGGGCTTTTCGGGTCAAACCAGGGTAGCTACGGCTCAT 2850
             |||||||||||||||||||||||||||||||||||||||||||
Db          42 CAAGGGGGCTTTTCGGGTCAAACCAGGGTAGCTACGGCTCAT 1

RESULT 3
US-09-134-246-4/c
; Sequence 4, Application US/09134246B
; Patent No. 6207377

; GENERAL INFORMATION:
;  APPLICANT: Wayne, Jay
;  APPLICANT: Xu, Shuang-yong
;  TITLE OF INVENTION: Method For Construction Of Thermus-E. coli Shuttle
;  TITLE OF INVENTION: Vectors And Identification of Two Thermus Plasmid
;  TITLE OF INVENTION: Replication Origins
;  FILE REFERENCE: Thermus Shuttle Vector
;  CURRENT APPLICATION NUMBER: US/09/134,246B
;  CURRENT FILING DATE: 1998-08-14
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 1026
;   TYPE: DNA
;   ORGANISM: Thermus sp.
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1023)
US-09-134-246-4

  Query Match            17.5%;  Score 1026;  DB 3;  Length 1026;
  Best Local Similarity  100.0%;  Pred. No. 9e-280;
  Matches 1026;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3132 TCAGGAGGCCTCCTTGAGGGTCTTCACTAGGTACGCTCCCGGTCGTGTCAGATGGCCATC 3191
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1026 TCAGGAGGCCTCCTTGAGGGTCTTCACTAGGTACGCTCCCGGTCGTGTCAGATGGCCATC 967

Qy       3192 CGTGTAATCTCGGATAACCCGCATTAGGACGGCAAATAGATATTGCGCGGGGAGTTCACC 3251
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        966 CGTGTAATCTCGGATAACCCGCATTAGGACGGCAAATAGATATTGCGCGGGGAGTTCACC 907

Qy       3252 CCTGGCCACAGCCCACAGCAAGCCTGCATAGAAACGTCTTGAACGACGGTCATCTAGGAG 3311
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        906 CCTGGCCACAGCCCACAGCAAGCCTGCATAGAAACGTCTTGAACGACGGTCATCTAGGAG 847

Qy       3312 ATCGGCAATGTACGTAGCAATGAGGGTGATAAGGGCCGGGAGTTTGGAACGCTCCACCTC 3371
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        846 ATCGGCAATGTACGTAGCAATGAGGGTGATAAGGGCCGGGAGTTTGGAACGCTCCACCTC 787

Qy       3372 GGGGAGGACCAGGATGAGGCCCAGGTCAACGGCCACGGTCTTGGTGTTGGGCATCACCCT 3431
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        786 GGGGAGGACCAGGATGAGGCCCAGGTCAACGGCCACGGTCTTGGTGTTGGGCATCACCCT 727

Qy       3432 TTTCCCCTGAGCCCAGAGGACCAGCACGTCCAGGGTGGGGCGGATTCCGTGGTCCTGGTA 3491
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        726 TTTCCCCTGAGCCCAGAGGACCAGCACGTCCAGGGTGGGGCGGATTCCGTGGTCCTGGTA 667

Qy       3492 GGCCTTGACCCAGTTGAAGGAGAGCACGCCGTTGGCCATGTCTAGGGCGAGGTTCCTCCA 3551
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        666 GGCCTTGACCCAGTTGAAGGAGAGCACGCCGTTGGCCATGTCTAGGGCGAGGTTCCTCCA 607

Qy       3552 GGGGTAGATGTAGTCGTCCAGGGTGAGCCTGGCTTTCCCTGGCCTCAGCCGGACGGCCCA 3611
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        606 GGGGTAGATGTAGTCGTCCAGGGTGAGCCTGGCTTTCCCTGGCCTCAGCCGGACGGCCCA 547

Qy       3612 AAGGGTGCCGATGGCCCGGCGCTCCCCGTTGACGGTTTGGTGAAGGACGTCGGTGGCCAC 3671
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        546 AAGGGTGCCGATGGCCCGGCGCTCCCCGTTGACGGTTTGGTGAAGGACGTCGGTGGCCAC 487

Qy       3672 CAGGCCCTTTTTCTCAAGGACCTTCTTCCAGGCGTGGACGGTCTGCCGGGTGACCCCCAG 3731
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        486 CAGGCCCTTTTTCTCAAGGACCTTCTTCCAGGCGTGGACGGTCTGCCGGGTGACCCCCAG 427

Qy       3732 GTTGAGGGCCAACATCTCCAGGGGGACCATGAAGACGACCGTCCCCCACCTTCCTCCAGAG 3791
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        426 GTTGAGGGCCAACATCTCCAGGGGGACCATGAAGACGACCGTCCCCCACCTTCCTCCAGAG 367

Qy       3792 CTCCCGGTTGCCGTAGGGGATGGTGGAGCGGGCAATCTCCTGGAGGAGTTCCAGAAGCTT 3851
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        366 CTCCCGGTTGCCGTAGGGGATGGTGGAGCGGGCAATCTCCTGGAGGAGTTCCAGAAGCTT 307

Qy       3852 CTGAGGCCCGTCCTGGACGGCTTGACATACATTCCCAACGGGGGGTTCAGCTGGTGGTAT 3911
```

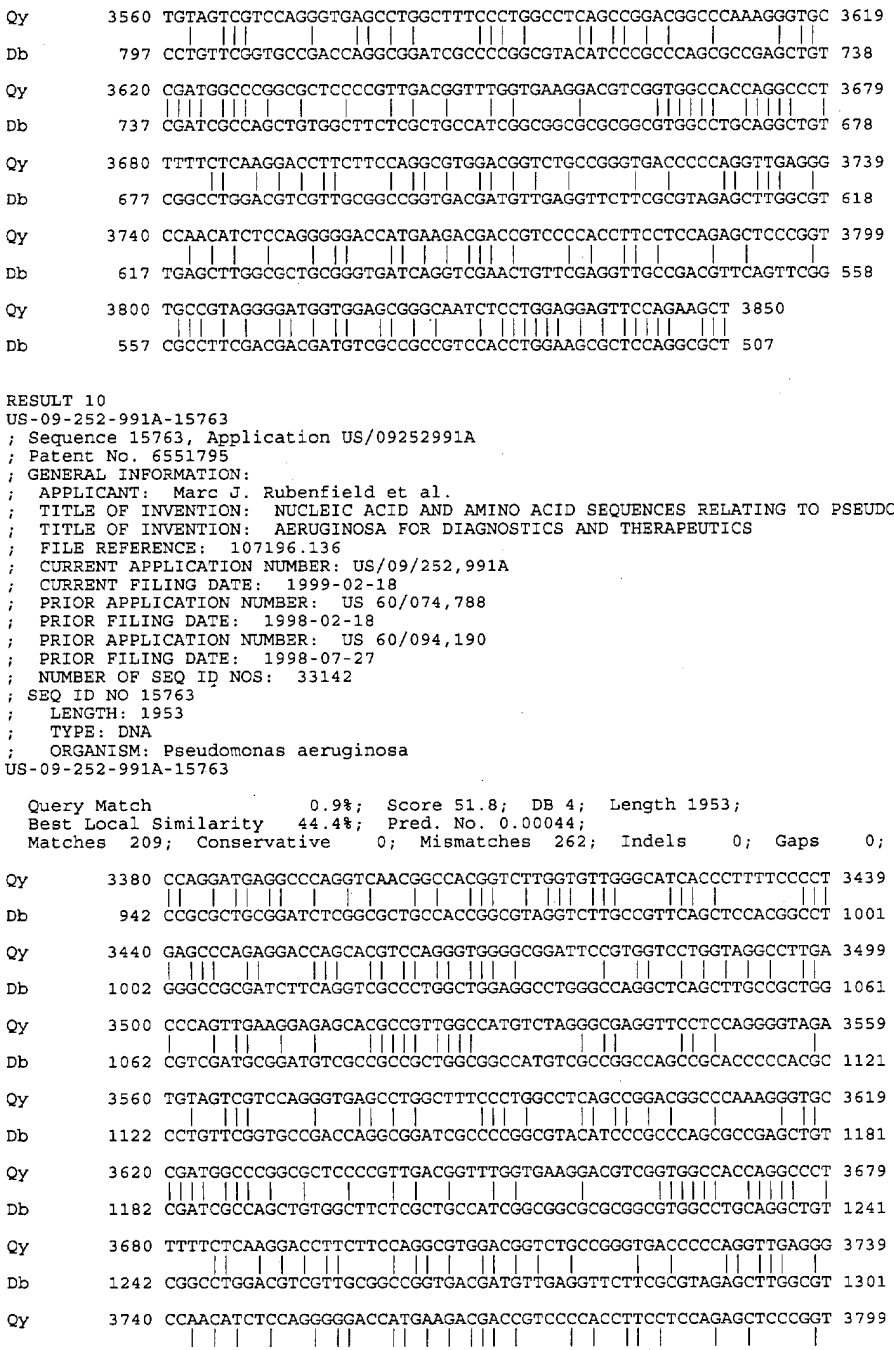
```
Db        306 CTGAGGCCCGTCCTGGACGGCTTGACATACATTCCCAACGGGGGGTTCAGCTGGTGGTAT 247

Qy       3912 AGGGGGGGCCCCAACAGAGGAAAGGAAAGCCTCATGGTCTTTCTCGGGTAAAGCAATCCT 3971

Db        246 AGGGGGGGCCCCAACAGAGGAAAGGAAAGCCTCATGGTCTTTCTCGGGTAAAGCAATCCT 187

Qy       3972 AAGTGCCTCTTTTTGGTATGTAAAGCCCTTCGCGAGGCGATTTTCGGCACCTCCATCTGG 4031

Db        186 AAGTGCCTCTTTTTGGTATGTAAAGCCCTTCGCGAGGCGATTTTCGGCACCTCCATCTGG 127

Qy       4032 AGGGGGGTCGGTGGCCAAGAAGAAGTCCTCTGACCCCCTATCTGACCCCCTAGTGGCATC 4091

Db        126 AGGGGGGTCGGTGGCCAAGAAGAAGTCCTCTGACCCCCTATCTGACCCCCTAGTGGCATC 67

Qy       4092 GGTGTTGTCGTGGGTTTCCTCTAAAGCCTCGTAAAGCTCTTCAAAGAAGGTTTTTTCGTT 4151

Db          66 GGTGTTGTCGTGGGTTTCCTCTAAAGCCTCGTAAAGCTCTTCAAAGAAGGTTTTTTCGTT 7

Qy       4152 CTTCAC 4157

Db           6 CTTCAC 1
```

```
RESULT 4
US-08-960-756-3
; Sequence 3, Application US/08960756
; Patent No. 5866422
;  GENERAL INFORMATION:
;    APPLICANT:  WAYNE, JAY
;    APPLICANT:  XU, SHUANG-YONG
;    TITLE OF INVENTION:  METHOD FOR CLONING AND
;    TITLE OF INVENTION:  PRODUCING THE Tsp45I RESTRICTION ENDONUCLEASE IN E. COLI
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  New England Biolabs, Inc.
;      STREET:  32 Tozer Road
;      CITY:  Beverly
;      STATE:  MA
;      COUNTRY:  US
;      ZIP:  01915
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/960,756
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Williams, Gregory D
;      REGISTRATION NUMBER:  30901
;      REFERENCE/DOCKET NUMBER:  NEB-128
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  978-927-5054
;      TELEFAX:  978-927-1705
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  999 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    FEATURE:
;      NAME/KEY:  Coding Sequence
;      LOCATION:  1...996
;      OTHER INFORMATION:
US-08-960-756-3
```

```
  Query Match             17.1%;  Score 999;  DB 2;  Length 999;
  Best Local Similarity   100.0%;  Pred. No. 3.8e-272;
  Matches  999;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        614 ATGCAACAGATGGCCGAGTGGAACGTGTGGACACAGAGAAGCGTTGAGCTTCTGGAGAAG 673

Db          1 ATGCAACAGATGGCCGAGTGGAACGTGTGGACACAGAGAAGCGTTGAGCTTCTGGAGAAG 60

Qy        674 GGGTATTTGGATAAACTACTGCAGGTCTATAAAGGGGAAAGTGGCTCTTCGAGGTCAGTA 733

Db         61 GGGTATTTGGATAAACTACTGCAGGTCTATAAAGGGGAAAGTGGCTCTTCGAGGTCAGTA 120

Qy        734 CCAGAGGAGGTAGAGGAAAAACTTCGCGAGGCCTACAAGGCATACGAGGGGAGGCAGGAT 793

Db        121 CCAGAGGAGGTAGAGGAAAAACTTCGCGAGGCCTACAAGGCATACGAGGGGAGGCAGGAT 180

Qy        794 AGTCCGGAGGCAGAAACGAAACTCGTGGAAGCCGTGCTAAATGCCAGAAAAAAGGTCGAG 853

Db        181 AGTCCGGAGGCAGAAACGAAACTCGTGGAAGCCGTGCTAAATGCCAGAAAAAAGGTCGAG 240

Qy        854 CGGTCCCCCTTCAATCACCCCTACCTGCCTTTGGTCTACTACCTGGTTTCGGAAAAAGCA 913

Db        241 CGGTCCCCCTTCAATCACCCCTACCTGCCTTTGGTCTACTACCTGGTTTCGGAAAAAGCA 300

Qy        914 GAAAAAGCGAACAAGGCCCTTGAGGAGGCATTGCAGGAGGTTGCCTCAAAGCACCCAGAA 973

Db        301 GAAAAAGCGAACAAGGCCCTTGAGGAGGCATTGCAGGAGGTTGCCTCAAAGCACCCAGAA 360

Qy        974 ACCATCCGCGTCCTGGCCAAGGAAGCGCAAAGAAGAGGCGTAGAAGCCTTGATCCAAAGG 1033

Db        361 ACCATCCGCGTCCTGGCCAAGGAAGCGCAAAGAAGAGGCGTAGAAGCCTTGATCCAAAGG 420

Qy       1034 CTCAAGGAGCCTCCCGAAATAAATCGGCAGATAGGGCCGATGTTCAAAAGGTGGTACAAA 1093

Db        421 CTCAAGGAGCCTCCCGAAATAAATCGGCAGATAGGGCCGATGTTCAAAAGGTGGTACAAA 480

Qy       1094 GAAGAGCTAAAGGGGAAAATAGAAGAGAGGCTTCCAGGCCCTACCAAACCAAAGATTGTG 1153

Db        481 GAAGAGCTAAAGGGGAAAATAGAAGAGAGGCTTCCAGGCCCTACCAAACCAAAGATTGTG 540

Qy       1154 GTAGTATCCCCTGAAAAAAGTAAACCGGAGCAAGCACCCCTTATTGCGGAGAGAGAAGCG 1213

Db        541 GTAGTATCCCCTGAAAAAAGTAAACCGGAGCAAGCACCCCTTATTGCGGAGAGAGAAGCG 600

Qy       1214 GGCATCATCATATACACGGGATCGGATGAAGCTTTGAAAGATGCCGCCAAGGAAAACCTG 1273

Db        601 GGCATCATCATATACACGGGATCGGATGAAGCTTTGAAAGATGCCGCCAAGGAAAACCTG 660

Qy       1274 GGCCTTGGCGAGGAAGCAGAACTAGGCACCAAGGGCGTAGATTTCTACGTGGTCATCCGG 1333

Db        661 GGCCTTGGCGAGGAAGCAGAACTAGGCACCAAGGGCGTAGATTTCTACGTGGTCATCCGG 720

Qy       1334 CGTAGCCCTGAAGAGACATGGCACCTAACAGGAGAAGTGAAGTTTCAATCCGACTTTGGC 1393

Db        721 CGTAGCCCTGAAGAGACATGGCACCTAACAGGAGAAGTGAAGTTTCAATCCGACTTTGGC 780

Qy       1394 GGAAACCAAGACAACCAGAAACTAGTAGCAAAGGCTTCCATAAGGTTGGACCTTGAGAAG 1453

Db        781 GGAAACCAAGACAACCAGAAACTAGTAGCAAAGGCTTCCATAAGGTTGGACCTTGAGAAG 840

Qy       1454 AGGCACATAGGAATAGTGGTGGTGGACGGAATGCCTGTGGTGAGCAAGTTTCGTGGGTGG 1513

Db        841 AGGCACATAGGAATAGTGGTGGTGGACGGAATGCCTGTGGTGAGCAAGTTTCGTGGGTGG 900

Qy       1514 GCCGGACTGGGGAAAGAAACGATCGTTACATCCGTACTCCTCCTTCCAGACCTGATAGCG 1573

Db        901 GCCGGACTGGGGAAAGAAACGATCGTTACATCCGTACTCCTCCTTCCAGACCTGATAGCG 960

Qy       1574 GAGCTCTACCAAAAGGGTGAAGAAGCCCTGGGCCTCTAG 1612

Db        961 GAGCTCTACCAAAAGGGTGAAGAAGCCCTGGGCCTCTAG 999
```

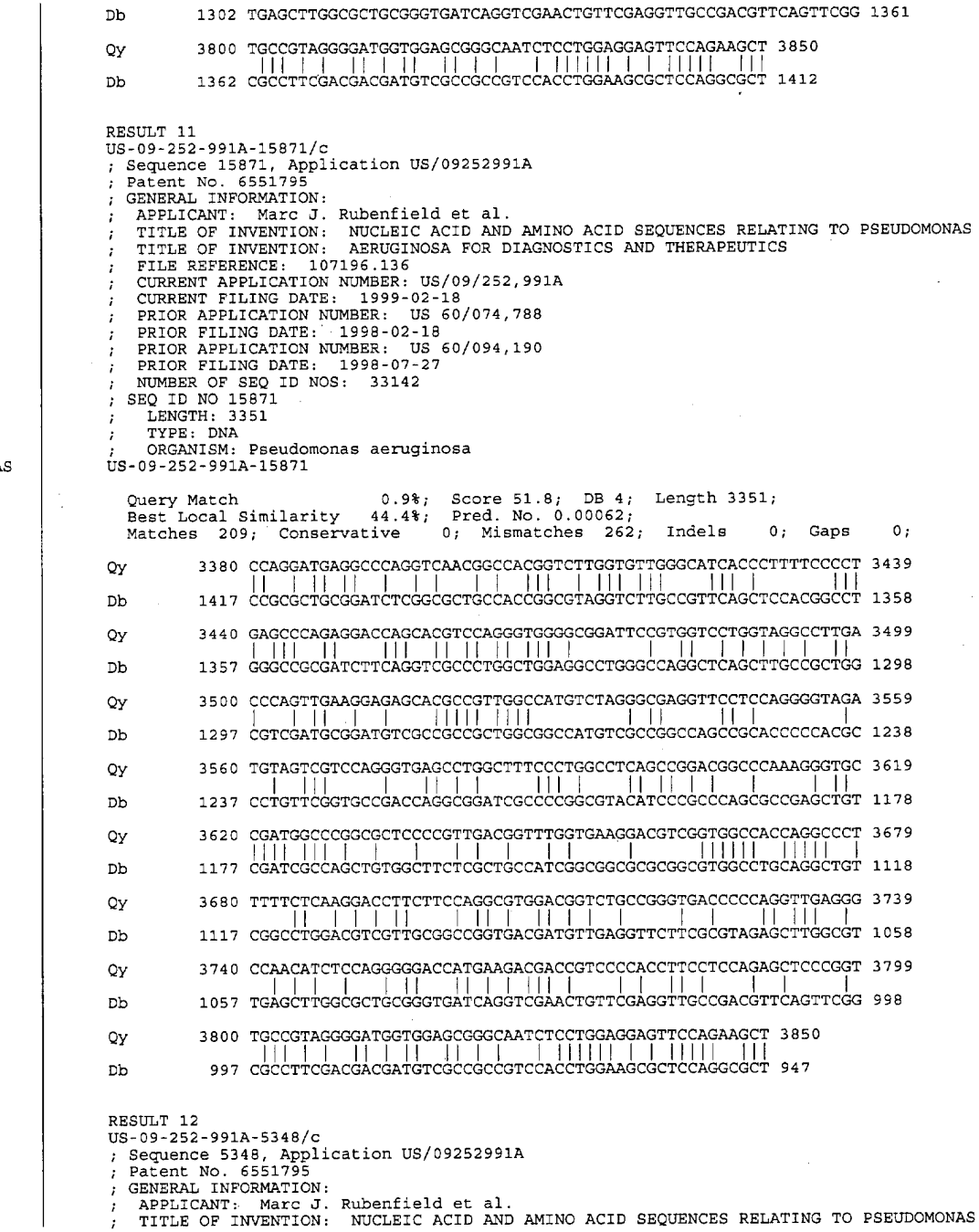
```
RESULT 5
US-09-134-246-5/c
; Sequence 5, Application US/09134246B
; Patent No. 6207377
; GENERAL INFORMATION:
;  APPLICANT: Wayne, Jay
;  APPLICANT: Xu, Shuang-yong
;  TITLE OF INVENTION: Method For Construction Of Thermus-E. coli Shuttle
;  TITLE OF INVENTION: Vectors And Identification of Two Thermus Plasmid
;  TITLE OF INVENTION: Replication Origins
;  FILE REFERENCE: Thermus Shuttle Vector
;  CURRENT APPLICATION NUMBER: US/09/134,246B
;  CURRENT FILING DATE: 1998-08-14
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 180
;   TYPE: DNA
;   ORGANISM: Thermus sp.
US-09-134-246-5

  Query Match             3.1%;  Score 180;  DB 3;  Length 180;
  Best Local Similarity   100.0%;  Pred. No. 5.5e-41;
  Matches  180;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4128 CTCTTCAAAGAAGGTTTTTTCGTTCTTCACCCTCGGACCTCCTTGTCATCTGGAGCCCGA 4187
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        180 CTCTTCAAAGAAGGTTTTTTCGTTCTTCACCCTCGGACCTCCTTGTCATCTGGAGCCCGA 121

Qy       4188 GGCGTTACCCTAGGTCTTGGGGGTGATCCGGGGCAACCGCCTCGGTTTCGCCTTTTTATG 4247
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        120 GGCGTTACCCTAGGTCTTGGGGGTGATCCGGGGCAACCGCCTCGGTTTCGCCTTTTTATG 61

Qy       4248 GGTCCAAAATAACCGTCAGCCCAGCGGCTGGCAATCCCCCCTCCTAAAAGGCCGTTATAG 4307
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          60 GGTCCAAAATAACCGTCAGCCCAGCGGCTGGCAATCCCCCCTCCTAAAAGGCCGTTATAG 1


RESULT 6
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14

  Query Match             1.3%;  Score 77;  DB 1;  Length 7218;
  Best Local Similarity   4.9%;  Pred. No. 7.5e-11;
  Matches   20;  Conservative  240;  Mismatches  145;  Indels    0;  Gaps    0;

Qy        469 GATTCCAGTATACCATAAATGGCACAAAGTTTTGAGAAGGTGGTCAAACAAAAAGGCTTT 528
              || |   ||| :  : ::: :: : ::::    ::::::: ::  ::: :::::::
Db       1445 GAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1386

Qy        529 CTCGGTCAGGTTATGGTGAGGTGGGGGCGGTCAAAGGCCGACTTAAGTTTGGTAAAGCCG 588
                ::  :::  : :: :::: ::::: ::  :::::  ::   :::   :: ::::  :
Db       1385 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1326

Qy        589 GGAGGAAGCAAACCGGGGTGTTACCATGCAACAGATGGCCGAGTGGAACGTGTGGACACA 648
              :::::::: :::  :::: :  :  : : :: ::: ::  ::: :::: : : ::: : :
Db       1325 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1266

Qy        649 GAGAAGCGTTGAGCTTCTGGAGAAGGGGTATTTGGATAAACTACTGCAGGTCTATAAAGG 708
              :::::: :  :::     :::::::::: :   ::: :::  :  : :::   : :::::
Db       1265 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1206

Qy        709 GGAAAGTGGCTCTTCGAGGTCAGTACCAGAGGAGGTAGAGGAAAAACTTCGCGAGGCCTA 768
              :::::: ::      ::::  :: :  :::::::: ::::::::::     : ::::   :
Db       1205 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1146

Qy        769 CAAGGCATACGAGGGGAGGCAGGATAGTCCGGAGGCAGAAACGAAACTCGTGGAAGCCGT 828
              :::: : : ::::::::: :::: ::   ::::: ::::: ::::   : :::::  :
Db       1145 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1086

Qy        829 GCTAAATGCCAGAAAAAAGGTCGAGCGGTCCCCCTTCAATCACCC 873
              :  ::: :  ::::::::: |||   | |||| |  |   || ||
Db       1085 RRRRRRRRRRRRRRRRRRATCGCAAGCTCCCTCGACCTGCAGCC 1041


RESULT 7
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
```

```
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14

  Query Match             1.0%;  Score 56.4;  DB 1;  Length 7218;
  Best Local Similarity   4.6%;  Pred. No. 5e-05;
  Matches   12;  Conservative  161;  Mismatches   87;  Indels    0;  Gaps    0;

Qy       4696 GGAGGCCGTTGTAGAGGGCCGTCTCGGGTTCAAATCCTTTCCCTCTCTCTCCAGGTTTCC 4755
Db       1045 GCAGGTCGAGGGAGCTTGCGATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1104

Qy       4756 GAGGTTCGAGGTCTTGGTCCAGGTCTTGTACCAAGTTTTGACCAAAGTCTATTCTCGGA 4815
Db       1105 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1164

Qy       4816 ATATAGGGGTATCTTGTCTATCTTCCCTACGGGATATCTCTGTCTGTGTGAACTTGATCC 4875
Db       1165 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1224

Qy       4876 CATCCCAATACATATCTCAATCTCCTAATCTCCTCTTCTCTCCAGATCCCTAATCTCTTC 4935
Db       1225 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1284

Qy       4936 TTCTACCTCTTTCTCCTCCC 4955
Db       1285 YYYYYYYYYYYYYYYYYYYY 1304

RESULT 8
US-09-252-991A-5309/c
; Sequence 5309, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 5309
;   LENGTH: 1551
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5309

  Query Match             0.9%;  Score 51.8;  DB 4;  Length 1551;
  Best Local Similarity   44.4%;  Pred. No. 0.00038;
  Matches  209;  Conservative    0;  Mismatches  262;  Indels    0;  Gaps    0;

Qy       3380 CCAGGATGAGGCCCAGGTCAACGGCCACGGTCTTGGTGTTGGGCATCACCCTTTTCCCCT 3439
Db        977 CCGCGCTGCGGATCTCGGCGCTGCCACCGGCGTAGGTCTTGCCGTTCAGCTCCACGGCCT 918

Qy       3440 GAGCCCAGAGGACCAGCACGTCCAGGGTGGGGCGGATTCCGTGGTCCTGGTAGGCCTTGA 3499
Db        917 GGGCCGCGATCTTCAGGTCGCCCTGGCTGGAGGCCTGGGCCAGGCTCAGCTTGCCGCTGG 858

Qy       3500 CCCAGTTGAAGGAGAGCACGCCGTTGGCCATGTCTAGGGCGAGGTTCCTCCAGGGGTAGA 3559
Db        857 CGTCGATGCGGATGTCGCCGCCGCTGGCGGCCATGTCGCCGGCCAGCCGCACCCCCACGC 798

Qy       3560 TGTAGTCGTCCAGGGTGAGCCTGGCTTTCCCTGGCCTCAGCCGGACGGCCCAAAGGGTGC 3619
Db        797 CCTGCTCGGTGCCGACCAGGCGGATCGCCCCGGCGTACATCCCGCCCAGCGCCGAGCTGT 738

Qy       3620 CGATGGCCCGGCGCTCCCCGTTGACGGTTTGGTGAAGGACGTCGGTGGCCACCAGGCCCT 3679
Db        737 CGATCGCCAGCTGTGGCTTCTCGCTGCCATCGGCGGCGCGCGGCGTGGCCTGCAGGCTGT 678

Qy       3680 TTTTCTCAAGGACCTTCTTCCAGGCGTGGACGGTCTGCCGGGTGACCCCCAGGTTGAGGG 3739
Db        677 CGGCCTGGACGTCGTTGCGGCCGGTGACGATGTTGAGGTTCTTCGCGTAGAGCTTGGCGT 618

Qy       3740 CCAACATCTCCAGGGGGACCATGAAGACGACCGTCCCCACCTTCCTCCAGAGCTCCCGGT 3799
Db        617 TGAGCTTGGCGCTGCGGGTGATCAGGTCGAACTGTTCGAGGTTGCCGACGTTCAGTTCGG 558

Qy       3800 TGCCGTAGGGGATGGTGGAGCGGGCAATCTCCTGGAGGAGTTCCAGAAGCT 3850
Db        557 CGCCTTCGACGACGATGTCGCCGCCGTCCACCTGGAAGCGCTCCAGGCGCT 507

RESULT 9
US-09-252-991A-15934/c
; Sequence 15934, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 15934
;   LENGTH: 1563
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-15934

  Query Match             0.9%;  Score 51.8;  DB 4;  Length 1563;
  Best Local Similarity   44.4%;  Pred. No. 0.00039;
  Matches  209;  Conservative    0;  Mismatches  262;  Indels    0;  Gaps    0;

Qy       3380 CCAGGATGAGGCCCAGGTCAACGGCCACGGTCTTGGTGTTGGGCATCACCCTTTTCCCCT 3439
Db        977 CCGCGCTGCGGATCTCGGCGCTGCCACCGGCGTAGGTCTTGCCGTTCAGCTCCACGGCCT 918

Qy       3440 GAGCCCAGAGGACCAGCACGTCCAGGGTGGGGCGGATTCCGTGGTCCTGGTAGGCCTTGA 3499
Db        917 GGGCCGCGATCTTCAGGTCGCCCTGGCTGGAGGCCTGGGCCAGGCTCAGCTTGCCGCTGG 858

Qy       3500 CCCAGTTGAAGGAGAGCACGCCGTTGGCCATGTCTAGGGCGAGGTTCCTCCAGGGGTAGA 3559
Db        857 CGTCGATGCGGATGTCGCCGCCGCTGGCGGCCATGTCGCCGGCCAGCCGCACCCCCACGC 798
```



RESULT 10
US-09-252-991A-15763
; Sequence 15763, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 15763
; LENGTH: 1953
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-15763

Query Match 0.9%; Score 51.8; DB 4; Length 1953;
Best Local Similarity 44.4%; Pred. No. 0.00044;
Matches 209; Conservative 0; Mismatches 262; Indels 0; Gaps 0;



RESULT 11
US-09-252-991A-15871/c
; Sequence 15871, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 15871
; LENGTH: 3351
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-15871

Query Match 0.9%; Score 51.8; DB 4; Length 3351;
Best Local Similarity 44.4%; Pred. No. 0.00062;
Matches 209; Conservative 0; Mismatches 262; Indels 0; Gaps 0;

RESULT 12
US-09-252-991A-5348/c
; Sequence 5348, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 5348
; LENGTH: 4158
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5348

Query Match 0.9%; Score 51.8; DB 4; Length 4158;
Best Local Similarity 44.4%; Pred. No. 0.00071;
Matches 209; Conservative 0; Mismatches 262; Indels 0; Gaps 0;

```
Qy       3380 CCAGGATGAGGCCCAGGTCAACGGCCACGGTCTTGGTGTTGGGCATCACCCTTTTCCCCT 3439
             ||  | ||  ||  |  | |  |  |  |||  | |||  |||     |||  |      |||
Db       1057 CCGCGCTGCGGATCTCGGCGCTGCCACCGGCGTAGGTCTTGCCGTTCAGCTCCACGGCCT 998

Qy       3440 GAGCCCAGAGGACCAGCACGTCCAGGGTGGGGCGGATTCCGTGGTCCTGGTAGGCCTTGA 3499
             | |||  ||    |||  || || || |||  |     |  ||  | | | | |  ||
Db        997 GGGCCGCGATCTTCAGGTCGCCCTGGCTGGAGGCCTGGGCCAGGCTCAGCTTGCCGCTGG 938

Qy       3500 CCCAGTTGAAGGAGAGCACGCCGTTGGCCATGTCTAGGGCGAGGTTCCTCCAGGGGTAGA 3559
             |   | ||  |  |    ||||| ||||          | ||     || |         |
Db        937 CGTCGATGCGGATGTCGCCGCCGCTGGCGGCCATGTCGCCGGCCAGCCGCACCCCCACGC 878

Qy       3560 TGTAGTCGTCCAGGGTGAGCCTGGCTTTCCCTGGCCTCAGCCGGACGGCCCAAAGGGTGC 3619
               |  |||     |  || | |     ||| |    || || | |  |     |  ||
Db        877 CCTGCTCGGTGCCGACCAGGCGGATCGCCCCGGCGTACATCCCGCCCAGCGCCGAGCTGT 818

Qy       3620 CGATGGCCCGGCGCTCCCCGTTGACGGTTTGGTGAAGGACGTCGGTGGCCACCAGGCCCT 3679
             ||||  |||  |  |    |  | |  |  |  |     |      ||||||  |||||  |
Db        817 CGATCGCCAGCTGTGGCTTCTCGCTGCCATCGGCGGCGCGCGGCGTGGCCTGCAGGCTGT 758

Qy       3680 TTTTCTCAAGGACCTTCTTCCAGGCGTGGACGGTCTGCCGGGTGACCCCCAGGTTGAGGG 3739
               ||  | | | ||    | || |  || | |  |     |  |    || |||  |
Db        757 CGGCCTGGACGTCGTTGCGGCCGGTGACGATGTTGAGGTTCTTCGCGTAGAGCTTGGCGT 698

Qy       3740 CCAACATCTCCAGGGGGACCATGAAGACGACCGTCCCCACCTTCCTCCAGAGCTCCCGGT 3799
               | | |  |   | |||   || | | ||| |    | |  || |    |  |     |
Db        697 TGAGCTTGGCGCTGCGGGTGATCAGGTCGAACTGTTCGAGGTTGCCGACGTTCAGTTCGG 638

Qy       3800 TGCCGTAGGGGATGGTGGAGCGGGCAATCTCCTGGAGGAGTTCCAGAAGCT 3850
               ||| | |  || | ||  || | |   | ||||| | | |||||  |||
Db        637 CGCCTTCGACGACGATGTCGCCGCCGTCCACCTGGAAGCGCTCCAGGCGCT 587
```

RESULT 13
US-09-252-991A-5227
; Sequence 5227, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 5227
; LENGTH: 4953
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5227

Query Match 0.9%; Score 51.8; DB 4; Length 4953;
Best Local Similarity 44.4%; Pred. No. 0.00079;
Matches 209; Conservative 0; Mismatches 262; Indels 0; Gaps 0;

```
Qy       3380 CCAGGATGAGGCCCAGGTCAACGGCCACGGTCTTGGTGTTGGGCATCACCCTTTTCCCCT 3439
             ||  | ||  ||  |  | |  |  |  |||  | |||  |||     |||  |      |||
Db       3942 CCGCGCTGCGGATCTCGGCGCTGCCACCGGCGTAGGTCTTGCCGTTCAGCTCCACGGCCT 4001

Qy       3440 GAGCCCAGAGGACCAGCACGTCCAGGGTGGGGCGGATTCCGTGGTCCTGGTAGGCCTTGA 3499
             | |||  ||    |||  || || || |||  |     |  ||  | | | | |  ||
Db       4002 GGGCCGCGATCTTCAGGTCGCCCTGGCTGGAGGCCTGGGCCAGGCTCAGCTTGCCGCTGG 4061

Qy       3500 CCCAGTTGAAGGAGAGCACGCCGTTGGCCATGTCTAGGGCGAGGTTCCTCCAGGGGTAGA 3559
             |   | ||  |  |    ||||| ||||          | ||     || |         |
Db       4062 CGTCGATGCGGATGTCGCCGCCGCTGGCGGCCATGTCGCCGGCCAGCCGCACCCCCACGC 4121

Qy       3560 TGTAGTCGTCCAGGGTGAGCCTGGCTTTCCCTGGCCTCAGCCGGACGGCCCAAAGGGTGC 3619
               |  |||     |  || | |     ||| |    || || | |  |     |  ||
Db       4122 CCTGCTCGGTGCCGACCAGGCGGATCGCCCCGGCGTACATCCCGCCCAGCGCCGAGCTGT 4181

Qy       3620 CGATGGCCCGGCGCTCCCCGTTGACGGTTTGGTGAAGGACGTCGGTGGCCACCAGGCCCT 3679
             ||||  |||  |  |    |  | |  |  |  |     |      ||||||  |||||  |
Db       4182 CGATCGCCAGCTGTGGCTTCTCGCTGCCATCGGCGGCGCGCGGCGTGGCCTGCAGGCTGT 4241

Qy       3680 TTTTCTCAAGGACCTTCTTCCAGGCGTGGACGGTCTGCCGGGTGACCCCCAGGTTGAGGG 3739
               ||  | | | ||    | || |  || | |  |     |  |    || |||  |
Db       4242 CGGCCTGGACGTCGTTGCGGCCGGTGACGATGTTGAGGTTCTTCGCGTAGAGCTTGGCGT 4301

Qy       3740 CCAACATCTCCAGGGGGACCATGAAGACGACCGTCCCCACCTTCCTCCAGAGCTCCCGGT 3799
               | | |  |   | |||   || | | ||| |    | |  || |    |  |     |
Db       4302 TGAGCTTGGCGCTGCGGGTGATCAGGTCGAACTGTTCGAGGTTGCCGACGTTCAGTTCGG 4361

Qy       3800 TGCCGTAGGGGATGGTGGAGCGGGCAATCTCCTGGAGGAGTTCCAGAAGCT 3850
               ||| | |  || | ||  || | |   | ||||| | | |||||  |||
Db       4362 CGCCTTCGACGACGATGTCGCCGCCGTCCACCTGGAAGCGCTCCAGGCGCT 4412
```

RESULT 14
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 0.8%; Score 49.6; DB 3; Length 4403765;
Best Local Similarity 45.7%; Pred. No. 0.23;
Matches 212; Conservative 0; Mismatches 249; Indels 3; Gaps 1;

```
Qy       3304 TCTAGGAGATCGGCAATGTACGTAGCAATGAGGGTGATAAGGGCCGGGAGTTTGGAACGC 3363
             ||  ||    | || | |  || | |   | | ||   ||  | |        | ||
Db    2780207 TCGTGGGTGACCGCCAGGATCGCACCGGGGTAGCTGGCCAGATGCTGTTCGAGCCACTGC 2780266

Qy       3364 TCCACCTCGGGGAGGACCAGGATGAGGCCCAGGTCAACGGCCACGGTCTTGGTGTTGGGC 3423
```

```
Db    2780267 ACACTTTCCGCGTCTAGGTGGTTGGTCGGCTCGTCGAGCAACAACAGGTCGGGTTTGGAC 2780326

Qy       3424 ATCACCCTTTTCCCCTGAGCCCAGAGGACCAGCACGTCCAGGGTGGGGCGGATTCCGTGG 3483

Db    2780327 AACAGCAGTTTGCACAGCGCCACCCGGCGACGCTCGCCACCGGATAGGTTGGTTACCGG- 2780385

Qy       3484 TCCTGGTAGGCCTTGACCCAGTTGAAGGAGAGCACGCCGTTGGCCATGTCTAGGGCGAGG 3543

Db    2780386 --CTCGTCGGCCGGCGGACAGCGCAGCGCATCCATGGCCTGCTCGAGCTGCGCGTCGAGG 2780443

Qy       3544 TTCCTCCAGGGGTAGATGTAGTCGTCCAGGGTGAGCCTGGCTTTCCCTGGCCTCAGCCGG 3603

Db    2780444 TCCCACGCGTCGGCGTGGTCCAGTTCCTCTTGCAGCCGACCCATCTCTTCCATCAGCTCG 2780503

Qy       3604 ACGGCCCAAAGGGTGCCGATGGCCCGGCGCTCCCCGTTGACGGTTTGGTGAAGGACGTCG 3663

Db    2780504 TCGGTGTAGTCGGTGGCCATCAATTCGGCGACCTCGTTGAAGCGGTCGAGCTTGATCTTG 2780563

Qy       3664 GTGGCCACCAGGCCCTTTTTCTCAAGGACCTTCTTCCAGGCGTGGACGGTCTGCCGGGTG 3723

Db    2780564 ATGTCCCCCATGCCCTCTTCCACATTGCCGCGAACGGTCTTGTCCTCGTTCAGCGGCGGT 2780623

Qy       3724 ACCCCCAGGTTGAGGGCCAACATCTCCAGGGGGACCATGAAGAC 3767

Db    2780624 TCCTGTTGCAGGATGCCCACGGTGGCGCCGGTGGCCAGGAAGGC 2780667
```

RESULT 15
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 0.8%; Score 49.6; DB 3; Length 4411529;
Best Local Similarity 45.7%; Pred. No. 0.23;
Matches 212; Conservative 0; Mismatches 249; Indels 3; Gaps 1;

```
Qy       3304 TCTAGGAGATCGGCAATGTACGTAGCAATGAGGGTGATAAGGGCCGGGAGTTTGGAACGC 3363

Db    2783397 TCGTGGGTGACCGCCAGGATCGCACCGGGGTAGCTGGCCAGATGCTGTTCGAGCCACTGC 2783456

Qy       3364 TCCACCTCGGGGAGGACCAGGATGAGGCCCAGGTCAACGGCCACGGTCTTGGTGTTGGGC 3423

Db    2783457 ACACTTTCCGCGTCTAGGTGGTTGGTCGGCTCGTCGAGCAACAACAGGTCGGGTTTGGAC 2783516

Qy       3424 ATCACCCTTTTCCCCTGAGCCCAGAGGACCAGCACGTCCAGGGTGGGGCGGATTCCGTGG 3483

Db    2783517 AACAGCAGTTTGCACAGCGCCACCCGGCGACGCTCGCCACCGGATAGGTTGGTTACCGG- 2783575

Qy       3484 TCCTGGTAGGCCTTGACCCAGTTGAAGGAGAGCACGCCGTTGGCCATGTCTAGGGCGAGG 3543

Db    2783576 --CTCGTCGGCCGGCGGACAGCGCAGCGCATCCATGGCCTGCTCGAGCTGCGCGTCGAGG 2783633

Qy       3544 TTCCTCCAGGGGTAGATGTAGTCGTCCAGGGTGAGCCTGGCTTTCCCTGGCCTCAGCCGG 3603

Db    2783634 TCCCACGCGTCGGCGTGGTCCAGTTCCTCTTGCAGCCGACCCATCTCTTCCATCAGCTCG 2783693

Qy       3604 ACGGCCCAAAGGGTGCCGATGGCCCGGCGCTCCCCGTTGACGGTTTGGTGAAGGACGTCG 3663

Db    2783694 TCGGTGTAGTCGGTGGCCATCAATTCGGCGACCTCGTTGAAGCGGTCGAGCTTGATCTTG 2783753

Qy       3664 GTGGCCACCAGGCCCTTTTTCTCAAGGACCTTCTTCCAGGCGTGGACGGTCTGCCGGGTG 3723

Db    2783754 ATGTCCCCCATGCCCTCTTCCACATTGCCGCGAACGGTCTTGTCCTCGTTCAGCGGCGGT 2783813

Qy       3724 ACCCCCAGGTTGAGGGCCAACATCTCCAGGGGGACCATGAAGAC 3767

Db    2783814 TCCTGTTGCAGGATGCCCACGGTGGCGCCGGTGGCCAGGAAGGC 2783857
```

Search completed: March 18, 2004, 05:51:35
Job time : 287.095 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 02:39:53 ; Search time 1345.06 Seconds
(without alignments)
16085.819 Million cell updates/sec

Title: US-09-664-186-6
Perfect score: 5849
Sequence: 1 tctagaaggtcagggtggac.....tgtagaggaccagaagaccc 5849

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2438257 seqs, 1849576744 residues

Total number of hits satisfying chosen parameters: 4876514

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	53.8	0.9	9025608	14	US-10-156-761-1	Sequence 1, Appli
c	2	53.2	0.9	1755	15	US-10-260-238-721	Sequence 721, App
c	3	51.6	0.9	1815	9	US-09-758-269-13	Sequence 13, Appl
c	4	50.8	0.9	1018	12	US-10-425-114-35941	Sequence 35941, A
c	5	50.8	0.9	1077	14	US-10-156-761-6573	Sequence 6573, Ap
c	6	50.8	0.9	9025608	14	US-10-156-761-1	Sequence 1, Appli
c	7	50	0.9	675	14	US-10-156-761-6076	Sequence 6076, Ap
c	8	49.8	0.9	1389	14	US-10-156-761-5431	Sequence 5431, Ap
c	9	49.6	0.8	1674	12	US-10-282-122A-26289	Sequence 26289, A
c	10	49.6	0.8	1677	12	US-10-282-122A-28540	Sequence 28540, A
c	11	49.4	0.8	898	12	US-10-425-114-32815	Sequence 32815, A
c	12	49.2	0.8	1803	15	US-10-369-493-42225	Sequence 42225, A
c	13	48.8	0.8	1045	12	US-10-425-114-26418	Sequence 26418, A
c	14	48.6	0.8	960	15	US-10-369-493-34210	Sequence 34210, A
c	15	48.6	0.8	9369	14	US-10-200-562-190	Sequence 190, App
c	16	48.6	0.8	9369	14	US-10-237-551-190	Sequence 190, App
c	17	48.6	0.8	9369	14	US-10-237-551-247	Sequence 247, App
	18	48.6	0.8	154746	10	US-09-827-688-8	Sequence 8, Appli
c	19	48	0.8	1977	14	US-10-355-430-17	Sequence 17, Appl
c	20	47.8	0.8	963	12	US-10-282-122A-11900	Sequence 11900, A
c	21	47.6	0.8	669	12	US-10-425-114-34873	Sequence 34873, A
c	22	47.6	0.8	864	12	US-10-425-114-35445	Sequence 35445, A
c	23	47.6	0.8	1527	14	US-10-156-761-6366	Sequence 6366, Ap
c	24	47.4	0.8	1926	14	US-10-156-761-2557	Sequence 2557, Ap
c	25	47.2	0.8	969	14	US-10-156-761-111	Sequence 111, App
c	26	46.8	0.8	1224	14	US-10-156-761-4891	Sequence 4891, Ap
c	27	46.8	0.8	1605	15	US-10-260-238-908	Sequence 908, App
c	28	46.8	0.8	2223	14	US-10-156-761-3605	Sequence 3605, Ap
c	29	46.6	0.8	1071	15	US-10-369-493-42984	Sequence 42984, A
c	30	46.6	0.8	10065	14	US-10-160-758-1	Sequence 1, Appli
c	31	46.6	0.8	10065	15	US-10-085-117-114	Sequence 114, App
c	32	46.6	0.8	10455	14	US-10-160-758-4	Sequence 4, Appli
c	33	46.6	0.8	11073	14	US-10-160-758-2	Sequence 2, Appli
c	34	46.6	0.8	11073	15	US-10-085-117-113	Sequence 113, App
c	35	46.4	0.8	1381	12	US-10-425-114-13514	Sequence 13514, A
c	36	46.2	0.8	1414	12	US-10-425-114-14100	Sequence 14100, A
	37	46.2	0.8	1473	9	US-09-853-161-43	Sequence 43, Appl
	38	46.2	0.8	1473	9	US-09-852-659A-43	Sequence 43, Appl
	39	46.2	0.8	1473	9	US-09-852-797-43	Sequence 43, Appl
c	40	46.2	0.8	2892	14	US-10-156-761-2448	Sequence 2448, Ap
c	41	46	0.8	808	12	US-10-424-599-52912	Sequence 52912, A
c	42	46	0.8	1263	15	US-10-369-493-40502	Sequence 40502, A
c	43	46	0.8	1418	12	US-10-425-114-1960	Sequence 1960, Ap
c	44	46	0.8	1532	12	US-10-425-114-34918	Sequence 34918, A
c	45	46	0.8	1864	12	US-10-425-114-28403	Sequence 28403, A

ALIGNMENTS

RESULT 1
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match 0.9%; Score 53.8; DB 14; Length 9025608;
Best Local Similarity 45.0%; Pred. No. 0.01;
Matches 202; Conservative 0; Mismatches 247; Indels 0; Gaps 0;

```
Qy       3449 GGACCAGCACGTCCAGGGTGGGGCGGATTCCGTGGTCCTGGTAGGCCTTGACCCAGTTGA 3508
             ||  | || ||  | |    | |||||     |  | |    | | | | ||||||| |||
Db    4922065 GGTGCTGCGCGGCGACCCAGAGGCGGTCCATGGCGCCGATCTCGTCGTCGACCCAGATGA 4922124
```

```
Qy       3509 AGGAGAGCACGCCGTTGGCCATGTCTAGGGCGAGGTTCCTCCAGGGGTAGATGTAGTCGT 3568
Db    4922125 ACGGGCGGTGGCCGGCCCACTCGACGAGATGGCGGGTCTTCCAGTGCAGGCCGCGCGGGC 4922184

Qy       3569 CCAGGGTGAGCCTGGCTTTCCCTGGCCTCAGCCGGACGGCCCAAAGGGTGCCGATGGCCC 3628
Db    4922185 CCTCGTGGGTGGAGGCTTCCGGCCACTCCACCACAGGCAGCCTGGGCAGTCCGATCCGTG 4922244

Qy       3629 GGCGCTCCCCGTTGACGGTTTGGTGAAGGACGTCGGTGGCCACCAGGCCCTTTTTCTCAA 3688
Db    4922245 GGGCGACGGACTCGTTGGCGTCATTCATCCAGGTCGAGGCCCACACCAGCTCACACCCCA 4922304

Qy       3689 GGACCTTCTTCCAGGCGTGGACGGTCTGCCGGGTGACCCCCAGGTTGAGGGCCAACATCT 3748
Db    4922305 GTGCCATCAGACGCGGTCCGACGGCCTGGTCGAGCCGCCCCAGCAGCGGGTTCCCCTCGT 4922364

Qy       3749 CCAGGGGGACCATGAAGACGACCGTCCCCACCTTCCTCCAGAGCTCCCGGTTGCCGTAGG 3808
Db    4922365 CGGAGAACGCCACGGAGCCGGCGGCACGGGGCTGGAGCCGCCGGGACGACGGCCCGAACG 4922424

Qy       3809 GGATGGTGGAGCGGGCAATCTCCTGGAGGAGTTCCAGAAGCTTCTGAGGCCCGTCCTGGA 3868
Db    4922425 GGATGAGGGGTCCGTCGACGTCGAGGAAAAGGACGGGCCGCTCCACAGAGCGGGTCATGA 4922484

Qy       3869 CGGCTTGACATACATTCCCAACGGGGGGT 3897
Db    4922485 CGGCACGATAACCCTGCGCGGGGACGGGT 4922513
```

RESULT 2
US-10-260-238-721/c
; Sequence 721, Application US/10260238
; Publication No. US20040016025A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, Paul R.
; APPLICANT: Moughamer, Todd G.
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiyaki
; APPLICANT: Kreps, Joel
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
; FILE REFERENCE: 60111-NP
; CURRENT APPLICATION NUMBER: US/10/260,238
; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/325,448
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 6077
; SEQ ID NO 721
; LENGTH: 1755
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: N_region
; LOCATION: (917)..(917)
; OTHER INFORMATION: n = any nucleotide
; FEATURE:
; NAME/KEY: N_region
; LOCATION: (1750)..(1750)
; OTHER INFORMATION: n = any nucleotide
US-10-260-238-721

Query Match 0.9%; Score 53.2; DB 15; Length 1755;
Best Local Similarity 50.8%; Pred. No. 0.00011;
Matches 127; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

```
Qy       3366 CACCTCGGGGAGGACCAGGATGAGGCCCAGGTCAACGGCCACGGTCTTGGTGTTGGGCAT 3425
Db        495 CACCTCGCCCTGGAGGACGAGGCTGTGGTGGTCGTCGTCGTGGGTGGAGTTCTTGGCGAT 436

Qy       3426 CACCCTTTTCCCCTGAGCCCAGAGGACCAGCACGTCCAGGGTGGGGCGGATTCCGTGGTC 3485
Db        435 CTGCTGCTTGACGTAGCCGACGGTGCGGCCGGGCTCGACGCGGAACCTGAACTTGTTGCC 376

Qy       3486 CTGGTAGGCCTTGACCCAGTTGAAGGAGAGCACGCCGTTGGCCATGTCTAGGGCGAGGTT 3545
Db        375 GTGGACGGTCTGGACGGTGATGAGGCGGAGGTCGGAGATGCGGATGACGAGGTGGAGGAC 316

Qy       3546 CCTCCAGGGGTAGATGTAGTCGTCCAGGGTGAGCCTGGCTTTCCCTGGCCTCAGCCGGAC 3605
Db        315 GTTGCCGTGGGAGACGCCGTAGTCCATGATGCGGCTGTCGTTCCGGGCGAGCTCCCGGCC 256

Qy       3606 GGCCCAAAGG 3615
Db        255 ATCGAAGACG 246
```

RESULT 3
US-09-758-269-13/c
; Sequence 13, Application US/09758269
; Patent No. US20020104120A1
; GENERAL INFORMATION:
; APPLICANT: IUCHI, SATOSHI
; APPLICANT: KOBAYASHI, MASATOMO
; APPLICANT: SHINOZAKI, KAZUO
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN
; TITLE OF INVENTION: CLEAVAGE ENZYME GENE
; FILE REFERENCE: 3914-3
; CURRENT APPLICATION NUMBER: US/09/758,269
; CURRENT FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: JP 2001-003476
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: JP 2000-010056
; PRIOR FILING DATE: 2000-01-13
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 1815
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1812)
US-09-758-269-13

Query Match 0.9%; Score 51.6; DB 9; Length 1815;
Best Local Similarity 45.9%; Pred. No. 0.00035;
Matches 177; Conservative 0; Mismatches 209; Indels 0; Gaps 0;

```
Qy       3453 CAGCACGTCCAGGGTGGGGCGGATTCCGTGGTCCTGGTAGGCCTTGACCCAGTTGAAGGA 3512
Db        540 CAGCGCGTGCACCATGCCGTCGCCGTCGAAGAGGTGGTGCCCCGCGACGGGGTCGAAGCA 481

Qy       3513 GAGCACGCCGTTGGCCATGTCTAGGGCGAGGTTCCTCCAGGGGTAGATGTAGTCGTCCAG 3572
Db        480 GGGGTTGGCGCCGTTGCGCGCGTAGACCCCGTCGATGAAGGGCGGGATGCGGCCGGAGAC 421

Qy       3573 GGTGAGCCTGGCTTTCCCTGGCCTCAGCCGGACGGCCCAAAGGGTGCCGATGGCCCGGCG 3632
Db        420 GGGGAGCTCGTGCACGGGCGGCCTCTCCCCGACGGGCGCGAAGTTGCCGGCGATCTGCAC 361

Qy       3633 CTCCCCGTTGACGGTTTGGTGAAGGACGTCGGTGGCCACCAGGCCCTTTTTCTCAAGGAC 3692
Db        360 GGCCGGGTCGGCCGTGCTGGGCAGCCCGTGGGGCCGCTCGAGGACGTTGGCCACGAACCC 301

Qy       3693 CTTCTTCCAGGCGTGGACGGTCTGCCGGGTGACCCCCAGGTTGAGGGCCAACATCTCCAG 3752
```

```
Db        300 TTCCTCGAACGCGTCGAGCGCGGCCGCCGCGGCGCGCTGGAACAAGTTGAGCTGCTTCTT 241

Qy       3753 GGGGACCATGAAGACGACCGTCCCCACCTTCCTCCAGAGCTCCCGGTTGCCGTAGGGGAT 3812

Db        240 GCCGCCCTCCGCTTTCCTCGGCGCCGCGGCGTGCCCTGGGACGGCAATGGCGGCGGGCTT 181

Qy       3813 GGTGGAGCGGGCAATCTCCTGGAGGA 3838

Db        180 CCTGGACGGCGCAGGCAGGTCGGCGA 155


RESULT 4
US-10-425-114-35941/c
; Sequence 35941, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 35941
;   LENGTH: 1018
;   TYPE: DNA
;   ORGANISM: Zea mays subsp. mexicana
;   FEATURE:
;   OTHER INFORMATION: Clone ID: UC-ZMROTEOSINTE020D09_FLI
US-10-425-114-35941

  Query Match             0.9%;  Score 50.8;  DB 12;  Length 1018;
  Best Local Similarity   46.6%;  Pred. No. 0.00043;
  Matches  163;  Conservative    0;  Mismatches  187;  Indels    0;  Gaps    0;

Qy       3525 GGCCATGTCTAGGGCGAGGTTCCTCCAGGGGTAGATGTAGTCGTCCAGGGTGAGCCTGGC 3584

Db        738 GTCCTTGTTGGTGTTGGTGTTAGTGTTGGTGTTGGCGTCGCCCTGCTTGTTGTCCCCGCC 679

Qy       3585 TTTCCCTGGCCTCAGCCGGACGGCCCAAAGGGTGCCGATGGCCCGGCGCTCCCCGTTGAC 3644

Db        678 CATCCCCAGCGTGCTCATCACCGCGTCCTTGGCGCCCGCCGCCGCGCTCTTCACCTGCTC 619

Qy       3645 GGTTTGGTGAAGGACGTCGGTGGCCACCAGGCCCTTTTTCTCAAGGACCTTCTTCCAGGC 3704

Db        618 AGTGGCCTGCTGGATGACGCCGCCGGACTTGTCCTTGCCGGAGACGGCGGTGTCCTTGGC 559

Qy       3705 GTGGACGGTCTGCCGGGTGACCCCCAGGTTGAGGGCCAACATCTCCAGGGGGACCATGAA 3764

Db        558 GTACTGCCCGGCCTCGGCGGCCTTCTGCTTGGCGGCCTCCATGGCGTCGGCTGCCTTCTG 499

Qy       3765 GACGACCGTCCCCACCTTCCTCCAGAGCTCCCGGTTGCCGTAGGGGATGGTGGAGCGGGC 3824

Db        498 CTTGGCGGCCTCCGTCGTCTCGCCGGCCTTCTGCTTGGCGGCCTCCGTCGTCTCGCCGGC 439

Qy       3825 AATCTCCTGGAGGAGTTCCAGAAGCTTCTGAGGCCCGTCCTGGACGGCTT 3874

Db        438 CTTCTGCTTGGCGGCCTCCGTCGTCTCGCCGGCCTTGTGCTTGGCGGCCT 389


RESULT 5
US-10-156-761-6573/c
; Sequence 6573, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
;  APPLICANT:  IKEDA, HARUO
;  APPLICANT:  ISHIKAWA, JUN
;  APPLICANT:  HORIKAWA, HIROSHI
;  APPLICANT:  SHIBA, TADAYOSHI
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  APPLICANT:  HATTORI, MASAHIRA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-262
;  CURRENT APPLICATION NUMBER: US/10/156,761
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 6573
;   LENGTH: 1077
;   TYPE: DNA
;   ORGANISM: Streptomyces avermitilis
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1077)
US-10-156-761-6573

  Query Match             0.9%;  Score 50.8;  DB 14;  Length 1077;
  Best Local Similarity   44.3%;  Pred. No. 0.00045;
  Matches  208;  Conservative    0;  Mismatches  262;  Indels    0;  Gaps    0;

Qy       3366 CACCTCGGGGAGGACCAGGATGAGGCCCAGGTCAACGGCCACGGTCTTGGTGTTGGGCAT 3425

Db       1029 CGCCGCCGGGATGCCGTCGAGGGCGATCCGCTCCGTCACCAGCGCGCCGAGGTCCAGGCG 970

Qy       3426 CACCCTTTTCCCCTGAGCCCAGAGGACCAGCACGTCCAGGGTGGGGCGGATTCCGTGGTC 3485

Db        969 CCCCGCCCGTACGTGCTCGGCCAGGACCGGCAGGTCCTTCGCCGGGTCCGAGTTGCCGTA 910

Qy       3486 CTGGTAGGCCTTGACCCAGTTGAAGGAGAGCACGCCGTTGGCCATGTCTAGGGCGAGGTT 3545

Db        909 CACGCAGCCCGACAGGGTGCGGCCCCAGTGGAAGATCTCCAGCGCGTTGAACACGACCTG 850

Qy       3546 CCTCCAGGGGTAGATGTAGTCGTCCAGGGTGAGCCTGGCTTTCCCTGGCCTCAGCCGGAC 3605

Db        849 CTGGTCCTTGCCGCCGATGCCGACGACCGTGGTGCGGCCGCCACGCCGCGTCGAGTCCCA 790

Qy       3606 GGCCCAAAGGGTGCCGATGGCCCGGCGCTCCCCGTTGACGGTTTGGTGAAGGACGTCGGT 3665

Db        789 GGCGGTGCGGATGGTCACGGCGCGGCCCACGCACTCGACGGCGACGTCCACCCCCTGCTT 730

Qy       3666 GGCCACCAGGCCCTTTTTCTCAAGGACCTTCTTCCAGGCGTGGACGGTCTGCCGGGTGAC 3725

Db        729 GTCGGTGAGCCCGCGGATCTCGCGGGCCGTGTTCTCGGAGGCGACGAGGTAGTCGGTGGC 670

Qy       3726 CCCCAGGTTGAGGGCCAACATCTCCAGGGGGACCATGAAGACGACCGTCCCCACCTTCCT 3785

Db        669 CCCGGCCCGGCGGGCGAGCGACTCCTTCTCCGGCGAGACGTCCACCGCGACGATCTTCGA 610

Qy       3786 CCAGAGCTCCCGGTTGCCGTAGGGGATGGTGGAGCGGGCAATCTCCTGGA 3835

Db        609 CGCGCCCGCGATCCGGGCGGCCTGGAGCGTCGCGAGGCCGACCCCGCCGA 560


RESULT 6
US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
;  APPLICANT:  IKEDA, HARUO
;  APPLICANT:  ISHIKAWA, JUN
;  APPLICANT:  HORIKAWA, HIROSHI
;  APPLICANT:  SHIBA, TADAYOSHI
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  APPLICANT:  HATTORI, MASAHIRA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-262
```

```
;  CURRENT APPLICATION NUMBER: US/10/156,761
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
;   LENGTH: 9025608
;   TYPE: DNA
;   ORGANISM: Streptomyces avermitilis
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4187715)
;   OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

  Query Match             0.9%;  Score 50.8;  DB 14;  Length 9025608;
  Best Local Similarity   44.3%;  Pred. No. 0.082;
  Matches  208;  Conservative    0;  Mismatches  262;  Indels    0;  Gaps    0;

Qy       3366 CACCTCGGGGAGGACCAGGATGAGGCCCAGGTCAACGGCCACGGTCTTGGTGTTGGGCAT 3425
Db    7895885 CGCCGCCGGGATGCCGTCGAGGGCGATCCGCTCCGTCACCAGCGCGCCGAGGTCCAGGCG 7895826

Qy       3426 CACCCTTTTCCCCTGAGCCCAGAGGACCAGCACGTCCAGGGTGGGGCGGATTCCGTGGTC 3485
Db    7895825 CCCCGCCCGTACGTGCTCGGCCAGGACCGGCAGGTCCTTCGCCGGGTCCGAGTTGCCGTA 7895766

Qy       3486 CTGGTAGGCCTTGACCCAGTTGAAGGAGAGCACGCCGTTGGCCATGTCTAGGGCGAGGTT 3545
Db    7895765 CACGCAGCCCGACAGGGTGCGGCCCCAGTGGAAGATCTCCAGCGCGTTGAACACGACCTG 7895706

Qy       3546 CCTCCAGGGGTAGATGTAGTCGTCCAGGGTGAGCCTGGCTTTCCCTGGCCTCAGCCGGAC 3605
Db    7895705 CTGGTCCTTGCCGCCGATGCCGACGACCGTGGTGCGGCCGCCACGCCGCGTCGAGTCCCA 7895646

Qy       3606 GGCCCAAAGGGTGCCGATGGCCCGGCGCTCCCCGTTGACGGTTTGGTGAAGGACGTCGGT 3665
Db    7895645 GGCGGTGCGGATGGTCACGGCGCGGCCCACGCACTCGACGGCGACGTCCACCCCCTGCTT 7895586

Qy       3666 GGCCACCAGGCCCTTTTTCTCAAGGACCTTCTTCCAGGCGTGGACGGTCTGCCGGGTGAC 3725
Db    7895585 GTCGGTGAGCCCGCGGATCTCGCGGGCCGTGTTCTCGGAGGCGACGAGGTAGTCGGTGGC 7895526

Qy       3726 CCCCAGGTTGAGGGCCAACATCTCCAGGGGGACCATGAAGACGACCGTCCCCACCTTCCT 3785
Db    7895525 CCCGGCCCGGCGGGCGAGCGACTCCTTCTCCGGCGAGACGTCCACCGCGACGATCTTCGA 7895466

Qy       3786 CCAGAGCTCCCGGTTGCCGTAGGGGATGGTGGAGCGGGCAATCTCCTGGA 3835
Db    7895465 CGCGCCCGCGATCCGGGCGGCCTGGAGCGTCGCGAGGCCGACCCCGCCGA 7895416


RESULT 7
US-10-156-761-6076/c
; Sequence 6076, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
;  APPLICANT:  IKEDA, HARUO
;  APPLICANT:  ISHIKAWA, JUN
;  APPLICANT:  HORIKAWA, HIROSHI
;  APPLICANT:  SHIBA, TADAYOSHI
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  APPLICANT:  HATTORI, MASAHIRA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-262
;  CURRENT APPLICATION NUMBER: US/10/156,761
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 6076
;   LENGTH: 675
;   TYPE: DNA
;   ORGANISM: Streptomyces avermitilis
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(675)
US-10-156-761-6076

  Query Match             0.9%;  Score 50;  DB 14;  Length 675;
  Best Local Similarity   44.7%;  Pred. No. 0.0006;
  Matches  237;  Conservative    0;  Mismatches  290;  Indels    3;  Gaps    1;

Qy       3241 GGGAGTTCACCCCTGGCCACAGCCCACAGCAAGCCTGCATAGAAACGTCTTGAACGACGG 3300
Db        632 GGGTCGTCCGCCTCGGTGATCGCGCGCACGACGACGACACGGCGGGCGCCCGCTTCGAGA 573

Qy       3301 TCATCTAGGAGATCGGCAATGTACGTAGCAATGAGGGTGATAAGGGCCGGGAGTTTGGAA 3360
Db        572 ACCTCGTCGAGATTGCCGAGGTCGATGCCGCCGATGGCGAACCAGGGGCGGTCGGTGCCC 513

Qy       3361 CGCTCCACCTCGGGGAGGACCAGGATGAGGCCCAGGTCAACGGCCACGGTCTTGGTGTTG 3420
Db        512 AGGGCGGCCGTGTGGCGGACCAGGTCGAGGCCCGGGGCGTGCCGGCCGGGCTTGGTGGGG 453

Qy       3421 GGCATCACCCTTTTCCCCTGAGCCCAGAGGACCAGCACGTCCAGGGTGGGGCGGATTCCG 3480
Db        452 GTGGGCCAGCAGGGACCCGTACAGAAGTAGTCCACGCCCTCCTGGACGGCGGCGGCCTCG 393

Qy       3481 TGGTCCTGGTAGGCCTTGACCCAGTTGAAGGAGAGCACGCCGTTGGCCATGTCTAGGGCG 3540
Db        392 GCCTCGGCCTCGGCATGCGTGGAGCGGCCGATGAGGACGTCGGCGCCGAGGATCGCCCGG 333

Qy       3541 AGGTTCCTCCAGGGGTAGATGTAGTCGTCCAGGGTGAGCCTGGCTTTCCCTGGCCTCAGC 3600
Db        332 GCCGCCGGGACGGGGAGGTCGCCCTGCCCCAGGTGGAGCACGTCGGAGTCGACGGCGTGC 273

Qy       3601 CGGACGGCCCAAAGGGTGCCGATGGCCCGGCGCTCCCCGTTGACGGTTTGGTGAAGGACG 3660
Db        272 GCGACGTCCGCCCGGTCGTTGACCGCGAGGAGCTTGCCGTGGCGG---CGACAGGCGTCC 216

Qy       3661 TCGGTGGCCACCAGGCCCTTTTTCTCAAGGACCTTCTTCCAGGCGTGGACGGTCTGCCGG 3720
Db        215 GCGAAGACCTGGAGGTGCTCCAGCTCCTCGGCCGCCTCCATGCCCTTGTCGCGCAGCTGC 156

Qy       3721 GTGACCCCCAGGTTGAGGGCCAACATCTCCAGGGGGACCATGAAGACGAC 3770
Db        155 ACGATGTCGACGCCGCCGGCCAGGGCCGCGTCGAGGAACTCGGCGAGGTC 106


RESULT 8
US-10-156-761-5431/c
; Sequence 5431, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
;  APPLICANT:  IKEDA, HARUO
;  APPLICANT:  ISHIKAWA, JUN
;  APPLICANT:  HORIKAWA, HIROSHI
;  APPLICANT:  SHIBA, TADAYOSHI
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  APPLICANT:  HATTORI, MASAHIRA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-262
;  CURRENT APPLICATION NUMBER: US/10/156,761
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
```

```
; SEQ ID NO 5431
;   LENGTH: 1389
;   TYPE: DNA
;   ORGANISM: Streptomyces avermitilis
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1389)
US-10-156-761-5431

  Query Match             0.9%;  Score 49.8;  DB 14;  Length 1389;
  Best Local Similarity   46.2%;  Pred. No. 0.001;
  Matches  165;  Conservative    0;  Mismatches  192;  Indels    0;  Gaps    0;

Qy       3364 TCCACCTCGGGGAGGACCAGGATGAGGCCCAGGTCAACGGCCACGGTCTTGGTGTTGGGC 3423
Db       1313 TCCTCGTCGTCCAGGTCGACGATCTCGCCGTTGGTGTCCTTCACCGTGGCGGCCTCGACC 1254

Qy       3424 ATCACCCTTTTCCCCTGAGCCCAGAGGACCAGCACGTCCAGGGTGGGGCGGATTCCGTGG 3483
Db       1253 ACGACGGCCAGCGCCTTGCCGCGGGCGACCTCGCCGACCAGCATCGGGACCTGGCCACCC 1194

Qy       3484 TCCTGGTAGGCCTTGACCCAGTTGAAGGAGAGCACGCCGTTGGCCATGTCTAGGGCGAGG 3543
Db       1193 TCGACGACCGCCTGGGCGAACTGGTCGGGGGACATGCCGGAGGAAGCCGCACGCCGCATG 1134

Qy       3544 TTCCTCCAGGGGTAGATGTAGTCGTCCAGGGTGAGCCTGGCTTTCCCTGGCCTCAGCCGG 3603
Db       1133 AGGTGCTCGGTGAGCTCCTCCTGGTTGACGTTGAGCTTCTCCTTGTTGACGAGCTCGTCG 1074

Qy       3604 ACGGCCCAAAGGGTGCCGATGGCCCGGCGCTCCCCGTTGACGGTTTGGTGAAGGACGTCG 3663
Db       1073 AGGACGAACTGCGTCTTGATGCCCTTGACCGCCGCTTCCTTGGTCTCGGCGTCGAACTCC 1014

Qy       3664 GTGGCCACCAGGCCCTTTTTCTCAAGGACCTTCTTCCAGGCGTGGACGGTCTGCCGG 3720
Db       1013 TCGACCGTCTTGCCCTGGATCTCGAGGTACTTCTCGAGGTCGAGGCCCATCTGGCCG 957

RESULT 9
US-10-282-122A-26289/c
; Sequence 26289, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 26289
;   LENGTH: 1674
;   TYPE: DNA
;   ORGANISM: Mycobacterium bovis
US-10-282-122A-26289

  Query Match             0.8%;  Score 49.6;  DB 12;  Length 1674;
  Best Local Similarity   45.7%;  Pred. No. 0.0013;
  Matches  212;  Conservative    0;  Mismatches  249;  Indels    3;  Gaps    1;

Qy       3304 TCTAGGAGATCGGCAATGTACGTAGCAATGAGGGTGATAAGGGCCGGGAGTTTGGAACGC 3363
Db        644 TCGTGGGTGACCGCCAGGATCGCACCGGGGTAGCTGGCCAGATGCTGTTCGAGCCACTGC 585

Qy       3364 TCCACCTCGGGGAGGACCAGGATGAGGCCCAGGTCAACGGCCACGGTCTTGGTGTTGGGC 3423
Db        584 ACACTTTCCGCGTCTAGGTGGTTGGTCGGCTCGTCGAGCAACAACAGGTCGGGTTTGGAC 525

Qy       3424 ATCACCCTTTTCCCCTGAGCCCAGAGGACCAGCACGTCCAGGGTGGGGCGGATTCCGTGG 3483
Db        524 AACAGCAGTTTGCACAGCGCCACCCGGCGACGCTCGCCACCGGATAGGTTGGTTACCGG- 466

Qy       3484 TCCTGGTAGGCCTTGACCCAGTTGAAGGAGAGCACGCCGTTGGCCATGTCTAGGGCGAGG 3543
Db        465 --CTCGTCGGCCGGCGGACAGCGCAGCGCATCCATGGCCTGCTCGAGCTGCGCGTCGAGG 408

Qy       3544 TTCCTCCAGGGGTAGATGTAGTCGTCCAGGGTGAGCCTGGCTTTCCCTGGCCTCAGCCGG 3603
Db        407 TCCCACGCGTCGGCGTGGTCCAGTTCCTCTTGCAGCCGACCCATCTCTTCCATCAGCTCG 348

Qy       3604 ACGGCCCAAAGGGTGCCGATGGCCCGGCGCTCCCCGTTGACGGTTTGGTGAAGGACGTCG 3663
Db        347 TCGGTGTAGTCGGTGGCCATCAATTCGGCGACCTCGTTGAAGCGGTCGAGCTTGATCTTG 288

Qy       3664 GTGGCCACCAGGCCCTTTTTCTCAAGGACCTTCTTCCAGGCGTGGACGGTCTGCCGGGTG 3723
Db        287 ATGTCCCCCATGCCCTCTTCCACATTGCCGCGAACGGTCTTGTCCTCGTTCAGCGGCGGT 228

Qy       3724 ACCCCCAGGTTGAGGGCCAACATCTCCAGGGGGACCATGAAGAC 3767
Db        227 TCCTGTTGCAGGATGCCCACGGTGGCGCCGGTGGCCAGGAAGGC 184

RESULT 10
US-10-282-122A-28540/c
; Sequence 28540, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
```

; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 28540
; LENGTH: 1677
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-10-282-122A-28540

Query Match 0.8%; Score 49.6; DB 12; Length 1677;
Best Local Similarity 45.7%; Pred. No. 0.0013;
Matches 212; Conservative 0; Mismatches 249; Indels 3; Gaps 1;

```
Qy       3304 TCTAGGAGATCGGCAATGTACGTAGCAATGAGGGTGATAAGGGCCGGGAGTTTGGAACGC 3363
Db        644 TCGTGGGTGACCGCCAGGATCGCACCGGGGTAGCTGGCCAGATGCTGTTCGAGCCACTGC 585

Qy       3364 TCCACCTCGGGGAGGACCAGGATGAGGCCCAGGTCAACGGCCACGGTCTTGGTGTTGGGC 3423
Db        584 ACACTTTCCGCGTCTAGGTGGTTGGTCGGCTCGTCGAGCAACAACAGGTCGGGTTTGGAC 525

Qy       3424 ATCACCCTTTTCCCCTGAGCCCAGAGGACCAGCACGTCCAGGGTGGGGCGGATTCCGTGG 3483
Db        524 AACAGCAGTTTGCACAGCGCCACCCGGCGACGCTCGCCACCGGATAGGTTGGTTACCGG- 466

Qy       3484 TCCTGGTAGGCCTTGACCCAGTTGAAGGAGAGCACGCCGTTGGCCATGTCTAGGGCGAGG 3543
Db        465 --CTCGTCGGCCGGCGGACAGCGCAGCGCATCCATGGCCTGCTCGAGCTGCGCGTCGAGG 408

Qy       3544 TTCCTCCAGGGGTAGATGTAGTCGTCCAGGGTGAGCCTGGCTTTCCCTGGCCTCAGCCGG 3603
Db        407 TCCCACGCGTCGGCGTGGTCCAGTTCCTCTTGCAGCCGACCCATCTCTTCCATCAGCTCG 348

Qy       3604 ACGGCCCAAAGGGTGCCGATGGCCCGGCGCTCCCCGTTGACGGTTTGGTGAAGGACGTCG 3663
Db        347 TCGGTGTAGTCGGTGGCCATCAATTCGGCGACCTCGTTGAAGCGGTCGAGCTTGATCTTG 288

Qy       3664 GTGGCCACCAGGCCCTTTTTCTCAAGGACCTTCTTCCAGGCGTGGACGGTCTGCCGGGTG 3723
Db        287 ATGTCCCCCATGCCCTCTTCCACATTGCCGCGAACGGTCTTGTCCTCGTTCAGCGGCGGT 228

Qy       3724 ACCCCCAGGTTGAGGGCCAACATCTCCAGGGGGACCATGAAGAC 3767
Db        227 TCCTGTTGCAGGATGCCCACGGTGGCGCCGGTGGCCAGGAAGGC 184
```

RESULT 11
US-10-425-114-32815/c
; Sequence 32815, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 32815
; LENGTH: 898
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLMO17024H02_FLI
US-10-425-114-32815

Query Match 0.8%; Score 49.4; DB 12; Length 898;
Best Local Similarity 44.2%; Pred. No. 0.0011;
Matches 203; Conservative 0; Mismatches 256; Indels 0; Gaps 0;

```
Qy       3370 TCGGGGAGGACCAGGATGAGGCCCAGGTCAACGGCCACGGTCTTGGTGTTGGGCATCACC 3429
Db        459 TAGGGGTCCTTAGGGATGCCCTCCATCCTGCCGGCCATGGTCACGTTCTCGGCGAACACG 400

Qy       3430 CTTTTCCCCTGAGCCCAGAGGACCAGCACGTCCAGGGTGGGGCGGATTCCGTGGTCCTGG 3489
Db        399 TCGCTGTAGTTGATGCCGGTGACCTCGGGCATGGCGTTGGGGTCGGACGTGTTGTCGGGG 340

Qy       3490 TAGGCCTTGACCCAGTTGAAGGAGAGCACGCCGTTGGCCATGTCTAGGGCGAGGTTCCTC 3549
Db        339 TGCTGCCCGTAGTTGCCGGTCATCCAGAACACCCACTTCATGGTGTGGAGGCTGAGGCCG 280

Qy       3550 CAGGGGTAGATGTAGTCGTCCAGGGTGAGCCTGGCTTTCCCTGGCCTCAGCCGGACGGCC 3609
Db        279 CGCACGAAGATGTCCTTGACGAAGCCGCCCCTCCCGGCACCGGACTTGACCCTGACGGCC 220

Qy       3610 CAAAGGGTGCCGATGGCCCGGCGCTCCCCGTTGACGGTTTGGTGAAGGACGTCGGTGGCC 3669
Db        219 GACTCCGTGTTGATGGCGACGCTGTCCTCGGCGCGCACGTCGCGGATGCCGCCGGACATC 160

Qy       3670 ACCAGGCCCTTTTTCTCAAGGACCTTCTTCCAGGCGTGGACGGTCTGCCGGGTGACCCCC 3729
Db        159 TCGCTGCCCAGCGCGATCATGGCGCTCGTGGGGGAGACGCAGGTCAGCCTCCGGATGACG 100

Qy       3730 AGGTTGAGGGCCAACATCTCCAGGGGGACCATGAAGACGACCGTCCCCACCTTCCTCCAG 3789
Db         99 ATGTGCTGGCTCGGCATGTTGAACCTGATGCCGTACTCGTCCCACCCGCTCTTCACGGCC 40

Qy       3790 AGCTCCCGGTTGCCGTAGGGGATGGTGGAGCGGGCAATC 3828
Db         39 ACGCAGTCGTCGCCGGAGACGATGTAGCAGTCCTCGATC 1
```

RESULT 12
US-10-369-493-42225/c
; Sequence 42225, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374

```
; SEQ ID NO 42225
;   LENGTH: 1803
;   TYPE: DNA
;   ORGANISM: Halobacterium sp. NRC-1
US-10-369-493-42225

  Query Match              0.8%;  Score 49.2;  DB 15;  Length 1803;
  Best Local Similarity   48.3%;  Pred. No. 0.0019;
  Matches  138;  Conservative    0;  Mismatches  148;  Indels    0;  Gaps    0;

Qy       3447 GAGGACCAGCACGTCCAGGGTGGGGCGGATTCCGTGGTCCTGGTAGGCCTTGACCCAGTT 3506
Db        483 GAGGATGTACGCGTAGTCGTCGGGCGTGAGCCGGTAGTTCAGTGGCGTGTGGATGGCCCC 424

Qy       3507 GAAGGAGAGCACGCCGTTGGCCATGTCTAGGGCGAGGTTCCTCCAGGGGTAGATGTAGTC 3566
Db        423 CAGCTGCATCGCGCCGTAGGCGGCTTCGAGGTGGTAGTGGGTGTTCGGGTCGAGGACGGC 364

Qy       3567 GTCCAGGGTGAGCCTGGCTTTCCCTGGCCTCAGCCGGACGGCCCAAAGGGTGCCGATGGC 3626
Db        363 GACGCGGTCGCCGCTGTCGATGCCGGCCGCCTGGAGGACCGCCGAAAAGCGGTCGGCGCG 304

Qy       3627 CCGGCGCTCCCCGTTGACGGTTTGGTGAAGGACGTCGGTGGCCACCAGGCCCTTTTTCTC 3686
Db        303 GTCGGCGAGCTCGCTGTACGTGTACCGCCGGCCGGTGGTGGCGACGACGGCGGTTTCGTC 244

Qy       3687 AAGGACCTTCTTCCAGGCGTGGACGGTCTGCCGGGTGACCCCCAGG 3732
Db        243 GCCGTAGTAGGTGCGCGCGCGGTCGATGAAGTCGGTGACGAGGAGG 198

RESULT 13
US-10-425-114-26418/c
; Sequence 26418, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 26418
;   LENGTH: 1045
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: LIB4172-028-B4_FLI
US-10-425-114-26418

  Query Match              0.8%;  Score 48.8;  DB 12;  Length 1045;
  Best Local Similarity   45.7%;  Pred. No. 0.0018;
  Matches  170;  Conservative    0;  Mismatches  202;  Indels    0;  Gaps    0;

Qy       3453 CAGCACGTCCAGGGTGGGGCGGATTCCGTGGTCCTGGTAGGCCTTGACCCAGTTGAAGGA 3512
Db        755 CAGCGCGTGCACCATGCCGTCGCCGTCGAAGAGGTGGTGCCCGGCGACGGGGTCGAAGCA 696

Qy       3513 GAGCACGCCGTTGGCCATGTCTAGGGCGAGGTTCCTCCAGGGGTAGATGTAGTCGTCCAG 3572
Db        695 GGGGTTGGCGCCGTTGCGCGCGTAGACCCCGTCGATGAAGGGCGGGATGCGGCCGGAGAC 636

Qy       3573 GGTGAGCCTGGCTTTCCCTGGCCTCAGCCGGACGGCCCAAAGGGTGCCGATGGCCCGGCG 3632
Db        635 GGGGAGCTCGTGCACGGGCGGCCTCTCCCCGACGGGCGCGAAGTTGCCGGCGATCTGCAC 576

Qy       3633 CTCCCCGTTGACGGTTTGGTGAAGGACGTCGGTGGCCACCAGGCCCTTTTTCTCAAGGAC 3692
Db        575 GGCCGGGTCGGCCGTGCTGGGCAGCCCGTGGGGCCGCTCCAGGACGTTGGCCACGAACCC 516

Qy       3693 CTTCTTCCAGGCGTGGACGGTCTGCCGGGTGACCCCCAGGTTGAGGGCCAACATCTCCAG 3752
Db        515 TTCCTCGAACGCGTCGAGCGCGGCCGCCGCGGCGCTGGAACAAGTTGAGCTGCTTCTT 456

Qy       3753 GGGGACCATGAAGACGACCGTCCCCACCTTCCTCCAGAGCTCCCGGTTGCCGTAGGGGAT 3812
Db        455 GCCGCCGTCCGCTTTCCTCGGCGCCGCGGAGTGCCCTGGGACGGCAATGGCGGCGGGCTT 396

Qy       3813 GGTGGAGCGGGC 3824
Db        395 CCTGGACGGCGC 384

RESULT 14
US-10-369-493-34210/c
; Sequence 34210, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
;  APPLICANT: Cao, Yongwei
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Chen, Xianfeng
;  TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;  TITLE OF INVENTION:  PLANTS WITH IMPROVED PROPERTIES
;  FILE REFERENCE: 38-10(52052)B
;  CURRENT APPLICATION NUMBER: US/10/369,493
;  CURRENT FILING DATE:  2003-02-28
;  PRIOR APPLICATION NUMBER: US 60/360,039
;  PRIOR FILING DATE: 2002-02-21
;  NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 34210
;   LENGTH: 960
;   TYPE: DNA
;   ORGANISM: Sphingomonas aromaticivorans
US-10-369-493-34210

  Query Match              0.8%;  Score 48.6;  DB 15;  Length 960;
  Best Local Similarity   45.0%;  Pred. No. 0.002;
  Matches  183;  Conservative    0;  Mismatches  224;  Indels    0;  Gaps    0;

Qy       3356 TGGAACGCTCCACCTCGGGGAGGACCAGGATGAGGCCCAGGTCAACGGCCACGGTCTTGG 3415
Db        610 TGGACAGGCCCATCTTGACGAGGTCGGCACCGGGATGCCGTTGACGGTCGAGTACTGGG 551

Qy       3416 TGTTGGGCATCACCCTTTTCCCCTGAGCCCAGAGGACCAGCACGTCCAGGGTGGGGCGGA 3475
Db        550 TGACCGGAACCATGGTGTCGCCGTGGCCGCCGAGAACGAACGTGTTCACGTCGCGGATCG 491

Qy       3476 TTCCGTGGTCCTGGTAGGCCTTGACCCAGTTGAAGGAGAGCACGCCGTTGGCCATGTCTA 3535
Db        490 AGACGCCGAATTCCCATGCCAGGAACGTGCTGAAGCGCGCCGAGTCGAGCACGCCGGCCA 431

Qy       3536 GGGCGAGGTTCCTCCAGGGGTAGATGTAGTCGTCCAGGGTGAGCCTGGCTTTCCCTGGCC 3595
Db        430 TGCCGACGACCTTGTTCGCGGGCAGGCCCGAGAACTCGCGCAGCGCCCAGACCATCGCGT 371

Qy       3596 TCAGCCGGACGGCCCAAAGGGTGCCGATGGCCCGGCGCTCCCCGTTGACGGTTTGGTGAA 3655
Db        370 CGAGCGGGTTGGTGATGCAGATCACGAACGCGTCGGGCGCGTTGTCGCGGATGCCTTCGC 311

Qy       3656 GGACGTCGGTGGCCACCAGGCCCTTTTTCTCAAGGACCTTCTTCCAGGCGTGGACGGTCT 3715
Db        310 CGACGGCCTTCATCACCTTGAGGTTGATGCCGAGCAGGTCGTCGCGGCTCATGCCGGGCT 251

Qy       3716 GCCGGGTGACCCCCAGGTTGAGGGCCAACATCTCCAGGGGGACCATG 3762
Db        250 TGCGGGCGACGCCGGCGGTGACGATGATCACGTCCGCGCCTGCGATG 204
```

```
RESULT 15
US-10-200-562-190/c
; Sequence 190, Application US/10200562
; Publication No. US20030165819A1
; GENERAL INFORMATION:
;  APPLICANT: McGowen, Patrick
;  APPLICANT:  Hosken, Nancy A.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
;  TITLE OF INVENTION:  TREATMENT OF HERPES SIMPLEX VIRUS INFECTION
;  FILE REFERENCE: 210121.538C2
;  CURRENT APPLICATION NUMBER: US/10/200,562
;  CURRENT FILING DATE:  2002-07-19
;  NUMBER OF SEQ ID NOS: 212
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 190
;   LENGTH: 9369
;   TYPE: DNA
;   ORGANISM: HSV2
US-10-200-562-190

  Query Match             0.8%;  Score 48.6;  DB 14;  Length 9369;
  Best Local Similarity   43.9%;  Pred. No. 0.0073;
  Matches  207;  Conservative    0;  Mismatches  264;  Indels    0;  Gaps    0;

Qy       3333 GAGGGTGATAAGGGCCGGGAGTTTGGAACGCTCCACCTCGGGGAGGACCAGGATGAGGCC 3392
              | |  ||   | | || || | ||| | |  | |    | |  ||  || | | | |||
Db       4113 GCGCTTGCGGAAGTCCCGGGGGTTGTAGCCGTGCGTGCCCGCCAGCGCCTGCAGGCGGCG 4054

Qy       3393 CAGGTCAACGGCCACGGTCTTGGTGTTGGGCATCACCCTTTTCCCCTGAGCCCAGAGGAC 3452
              ||| || ||  |  ||  || || |  | |   || |  | |  |   |||  || | |
Db       4053 CAGCTCGACCACGTCGAACTCGGCGCGGTTCTCGACGCGGTCCAGCGCCGCCTCGACGCC 3994

Qy       3453 CAGCACGTCCAGGGTGGGGCGGATTCCGTGGTCCTGGTAGGCCTTGACCCAGTTGAAGGA 3512
                |    |  | || |  | |  |  ||||  ||| |   |    | | |
Db       3993 GGCGGCCCAGCGCTCGCTGCTGCCCCGGGCGCGCTGGGCCGCCATCTTCGCCGTCAGGTC 3934

Qy       3513 GAGCACGCCGTTGGCCATGTCTAGGGCGAGGTTCCTCCAGGGGTAGATGTAGTCGTCCAG 3572
              |  ||| ||    | |  ||   |||| ||   | |  || |  ||||   | | ||||
Db       3933 GGCGACGGCGGCCTCAAGTTCGTCGGCGCGGCGTCGCGTGGCGCCGATGACCTTGCCCAG 3874

Qy       3573 GGTGAGCCTGGCTTTCCCTGGCCTCAGCCGGACGGCCCAAAGGGTGCCGATGGCCCGGCG 3632
                ||  |||   |||             || | |             ||  |  | | |
Db       3873 CTCCTGCAGGGCGCGCCCGCTGGGGGAATGGTCCCCGGCCGTCCCTTCGGCGTGCAGCAG 3814

Qy       3633 CTCCCCGTTGACGGTTTGGTGAAGGACGTCGGTGGCCACCAGGCCCTTTTTCTCAAGGAC 3692
               |||||   | |  | |||    ||  | |  ||       |  | || |  || |
Db       3813 GCCCCCGAACCCAGCCTCGTGCCCCGCGAGGCTTTCCCGAGCAGCGGTCGTCGCGCGGGC 3754

Qy       3693 CTTCTTCCAGGCGTGGACGGTCTGCCGGGTGACCCCCAGGTTGAGGGCCAACATCTCCAG 3752
              |     |  | || |||  ||        ||  | ||  |  ||||    | || |
Db       3753 CGCGGCATCGATGAGGGCGGCATGGTCCCCCTCCGGCTGGGCGCAGGCCCGGCGCGCCTG 3694

Qy       3753 GGGGACCATGAAGACGACCGTCCCCACCTTCCTCCAGAGCTCCCGGTTGCC 3803
              |  |||| |  | || ||| |  | ||    ||  ||||||   | || |
Db       3693 GACTACCAGGTCGGCGGCCGCCGACCCCAGGGTCGTGAGCTCGTCGATGGC 3643


Search completed: March 18, 2004, 10:21:56
Job time : 1368.06 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: March 17, 2004, 21:14:07 ; Search time 9459.64 Seconds
(without alignments)
18464.115 Million cell updates/sec

Title: US-09-664-186-6
Perfect score: 5849
Sequence: 1 tctagaaggtcagggtggac.....tgtagaggaccagaagaccc 5849

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	60	1.0	683	14	CA235176	CA235176 SCCCFL409
c	2	58.6	1.0	997	29	CNS005TE	AL060767 Drosophil
	3	54.6	0.9	925	29	CNS0091P	AL053013 Drosophil
c	4	53.2	0.9	655	29	CG352167	CG352167 OG1BW12TV
	5	53.2	0.9	679	29	CC609109	CC609109 OGWDJ44TH
	6	53.2	0.9	773	29	CC661222	CC661222 OGLAU18TV
c	7	53.2	0.9	807	28	BZ529007	BZ529007 OGAIR91TC
c	8	53.2	0.9	815	29	CG034209	CG034209 PUJGY50TD
c	9	53.2	0.9	897	29	CC661216	CC661216 OGLAU18TH
	10	52.4	0.9	1201	13	BX381961	BX381961 BX381961
c	11	51.8	0.9	925	29	CNS0091P	AL053013 Drosophil
c	12	51.6	0.9	497	14	CA552474	CA552474 C0857F07-
c	13	51.2	0.9	649	14	CA230834	CA230834 SCJFFL3C0
c	14	51.2	0.9	649	14	CB870888	CB870888 HC15J04w
	15	50.6	0.9	594	14	CB874255	CB874255 HC15J04y
c	16	50.6	0.9	637	12	BI960118	BI960118 HVSMEn002
c	17	50.6	0.9	664	13	BQ762267	BQ762267 EBrc01_SQ
c	18	50.6	0.9	973	14	CA275829	CA275829 SCCCSD109
c	19	50.4	0.9	855	29	CG041883	CG041883 PUJEF24TD
c	20	50.4	0.9	890	29	CC731258	CC731258 OGLAX50TH
c	21	50.4	0.9	937	29	CC729649	CC729649 OGLAJ02TV
	22	50.4	0.9	1201	13	BX376097	BX376097 BX376097
c	23	50.2	0.9	509	12	BJ549094	BJ549094 BJ549094
c	24	49.8	0.9	608	14	CA179624	CA179624 SCVPST106
c	25	49.8	0.9	659	14	CA194749	CA194749 SCRFSB102
c	26	49.8	0.9	694	14	CA196087	CA196087 SCSBAD108
c	27	49.8	0.9	1201	13	BX381961	BX381961 BX381961
c	28	49.4	0.8	642	29	CG043714	CG043714 PUILF82TB
	29	49.4	0.8	775	29	CG207191	CG207191 OGXFN36TH
c	30	49.4	0.8	793	29	CG449321	CG449321 OGVHG34TH
	31	49.4	0.8	837	29	CC701337	CC701337 OGUKL21TV
	32	49.2	0.8	502	12	BM335757	BM335757 MEST160-F
	33	49.2	0.8	728	14	CF638161	CF638161 zmrww00_0
c	34	49.2	0.8	759	10	BF259495	BF259495 HVSMEf001
c	35	49	0.8	745	29	CC633883	CC633883 OGNAK47TV
	36	49	0.8	747	29	CC633878	CC633878 OGNAK47TH
c	37	49	0.8	869	29	CG284200	CG284200 OG0BZ21TV
	38	49	0.8	909	29	CG228851	CG228851 OG2BX92TH
c	39	49	0.8	1201	13	BX376097	BX376097 BX376097
c	40	48.6	0.8	559	14	CA208502	CA208502 SCMCSB111
c	41	48.6	0.8	564	14	CA210811	CA210811 SCEPSB113
c	42	48.6	0.8	605	14	CA181598	CA181598 SCBFST313
c	43	48.6	0.8	620	14	CA182549	CA182549 SCCCST308
c	44	48.6	0.8	647	14	CA216848	CA216848 SCRFAD112
c	45	48.6	0.8	730	14	CA195888	CA195888 SCEZAD108

ALIGNMENTS

RESULT 1
CA235176/c
LOCUS CA235176 683 bp mRNA linear EST 25-SEP-2003
DEFINITION SCCCFL4093A04.g FL4 Saccharum officinarum cDNA clone SCCCFL4093A04 5', mRNA sequence.
ACCESSION CA235176
VERSION CA235176.1 GI:35303456
KEYWORDS EST.
SOURCE Saccharum officinarum
ORGANISM Saccharum officinarum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Saccharum.
REFERENCE 1 (bases 1 to 683)
AUTHORS Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.
TITLE The libraries that made SUCEST
JOURNAL Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
COMMENT Contact: Arruda P
Centro de Biologia Molecular e Engenharia Genetica
Universidade Estadual de Campinas
Caixa Postal 6010, 13083-970, Campinas SP, Brazil
Tel: 55 19 3788 1137
Fax: 55 19 3788 1089
Email: parruda@unicamp.br
Clone distribution: clone distribution information can be found through the Brazilian Clone Collection Center (BCCC) at http://www.bcccenter.fcav.unesp.br

Plate: 093 row: A column: 04
Seq primer: T7 Promoter Primer.

FEATURES Location/Qualifiers
source 1. .683
/organism="Saccharum officinarum"
/mol_type="mRNA"
/db_xref="taxon:4547"
/clone="SCCCFL4093A04"
/lab_host="DH10B"
/clone_lib="FL4"
/note="Organ: Developed inflorescence and rachis (20cm-long); Vector: pSport1; Site_1: SalI; Site_2: NotI; An unidirectional cDNA library generated from [Developed inflorescence and rachis (20cm-long)]. cDNA was prepared from polyA+ mRNA using SuperScript Plasmid System Kit (Invitrogen). The double-strand cDNAs were fractionated in a sepharose CL-2B 40cm-columns and fragments sizing between 0.8 and 1.5 Kb were directionally cloned into the vector. Details of each source of RNA and library construction can be obtained at http://sucest.lad.ic.unicamp.br/public"

ORIGIN

Query Match 1.0%; Score 60; DB 14; Length 683;
Best Local Similarity 46.7%; Pred. No. 0.0067;
Matches 189; Conservative 0; Mismatches 216; Indels 0; Gaps 0;

```
Qy       3397 TCAACGGCCACGGTCTTGGTGTTGGGCATCACCCTTTTCCCCTGAGCCCAGAGGACCAGC 3456
Db        658 TGAAAGGGCCAGGGCCAGGTGCGGGNCCCCGTTCCTTCCCCCCGACTTGGGCGGTGGGGC 599

Qy       3457 ACGTCCAGGGTGGGGCGGATTCCGTGGTCCTGGTAGGCCTTGACCCAGTTGAAGGAGAGC 3516
Db        598 GCATCGGGTTGCCGCCGCACTCCCTTGTCAACGTAGACGTAGTACCTGAGCACGAAGGCG 539

Qy       3517 ACGCCGTTGGCCATGTCTAGGGCGAGGTTCCTCCAGGGGTAGATGTAGTCGTCCAGGGTG 3576
Db        538 AGCGCGAGCGCGACCCACTCGAGGCAGAAGATGGCGACGCCGAGCCCGCCGACCATGCGG 479

Qy       3577 AGCCTGGCTTTCCCTGGCCTCAGCCGGACGGCCCAAAGGGTGCCGATGGCCCGGCGCTCC 3636
Db        478 AGGATGACGGCGCCGTCCTCCTCGCGGACGTACGACCTGAGCCCGTCGGAGAGGAACTCC 419

Qy       3637 CCGTTGACGGTTTGGTGAAGGACGTCGGTGGCCACCAGGCCCTTTTTCTCAAGGACCTTC 3696
Db        418 GCGGTGCGGGAGAAGGCGAGCACGGCGGCGGCGCCCTGCAGGATGGAGACGACGACGGTG 359

Qy       3697 TTCCAGGCGTGGACGGTCTGCCGGGTGACCCCCAGGTTGAGGGCCAACATCTCCAGGGGG 3756
Db        358 GCCGACATGTGGACGGTGTAGCAGCGGCCCCGCGGCGGGTCGGGCGCCGCGATGACGGCG 299

Qy       3757 ACCATGAAGACGACCGTCCCCACCTTCCTCCAGAGCTCCCGGTTG 3801
Db        298 CACCCGGCCACGGCGGCGGCCACGGTCAGCGCGTGGAGCAGGATG 254
```

RESULT 2
CNS005TE/c
LOCUS CNS005TE 997 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR12K22 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL060767
VERSION AL060767.1 GI:4943573
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 997)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES Location/Qualifiers
source 1. .997
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR12K22"
/clone_lib="RPCI-98"
/note="end : TET3"

ORIGIN

Query Match 1.0%; Score 58.6; DB 29; Length 997;
Best Local Similarity 18.7%; Pred. No. 0.018;
Matches 98; Conservative 180; Mismatches 245; Indels 0; Gaps 0;

```
Qy        905 GAAAAAGCAGAAAAAGCGAACAAGGCCCTTGAGGAGGCATTGCAGGAGGTTGCCTCAAAG 964
Db        989 GARARARARRGARRRRGRRGARGARARRARRRRARGARAGARAARAGARRAGAGARARAA 930

Qy        965 CACCCAGAAACCATCCGCGTCCTGGCCAAGGAAGCGCAAAGAAGAGGCGTAGAAGCCTTG 1024
Db        929 ARRRARRRRAGRRAGRAGRGRSAGGGRGRGGRRAAAARAARDARARARARARARRRARAK 870

Qy       1025 ATCCAAAGGCTCAAGGAGCCTCCCGAAATAAATCGGCAGATAGGGCCGATGTTCAAAAGG 1084
Db        869 AARARGARRRRGRRRRRGRAGRRRARGAGGRAGRRGRGRGRRRARARRRAGAARARARGR 810

Qy       1085 TGGTACAAAGAAGAGCTAAAGGGGAAAATAGAAGAGAGGCTTCCAGGCCCTACCAAACCA 1144
Db        809 RARAARARARRRRAGRRRRRARGARGRRRRGGRRRGRGGRRRGRRRRAARRARARARAGA 750

Qy       1145 AAGATTGTGGTAGTATCCCCTGAAAAAAGTAAACCGGAGCAAGCACCCCTTATTGCGGAG 1204
Db        749 RAAARRRRARRAGAAARRARRRRRRRRRRGAGARARRARRRRARRAARAGAGAARRRRGRR 690

Qy       1205 AGAGAAGCGGGCATCATCATATACACGGGATCGGATGAAGCTTTGAAAGATGCCGCCAAG 1264
Db        689 GRGAGARRRGRGRGRRRGAGARRRRRGRRMRTRRARRRRRARRGAGAARRRARGARRRRR 630

Qy       1265 GAAAACCTGGGCCTTGGCGAGGAAGCAGAACTAGGCACCAAGGGCGTAGATTTCTACGTG 1324
Db        629 GRRGRRRRGRARGRRRRAGRGRRRGRRARGARRGRRRAGRRRRRRARAGAGARRGRRRRG 570

Qy       1325 GTCATCCGGCGTAGCCCTGAAGAGACATGGCACCTAACAGGAGAAGTGAAGTTTCAATCC 1384
Db        569 RARRAARGAARRRRARRARRRRRRAAARRGAGRAGARARGRAGRARRARRARRRRRRRG 510

Qy       1385 GACTTTGGCGGAAACCAAGACAACCAGAAACTAGTAGCAAAGG 1427
Db        509 GARRRGGRRGGRRRAAGARRRRAARAAARARRRARAGRAGARR 467
```

RESULT 3
CNS0091P
LOCUS CNS0091P 925 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #

```
            BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL053013
VERSION     AL053013.1  GI:4934461
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 925)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .925
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR19D16"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match             0.9%;  Score 54.6;  DB 29;  Length 925;
  Best Local Similarity   15.8%;  Pred. No. 0.16;
  Matches   64;  Conservative  166;  Mismatches  176;  Indels    0;  Gaps    0;

Qy       3499 ACCCAGTTGAAGGAGAGCACGCCGTTGGCCATGTCTAGGGCGAGGTTCCTCCAGGGGTAG 3558
Db        520 ANANANNNNNAGCSMSMCGKKCGSTTBGSTTTTTSSGSGYGKGCSSGSGBSCSCCSSCS 579

Qy       3559 ATGTAGTCGTCCAGGGTGAGCCTGGCTTTCCCTGGCCTCAGCCGGACGGCCCAAAGGGTG 3618
Db        580 CSSSSCSCCBCCCCCSCSSYCCSSSBSSSKCSSTSBSCSCCCSSKSVCGTSCSSSSSCSS 639

Qy       3619 CCGATGGCCCGGCGCTCCCCGTTGACGGTTTGGTGAAGGACGTCGGTGGCCACCAGGCCC 3678
Db        640 SSSSTSSSSTSSSTSKSSSSGSSSSSSSSYTTSKSTSASGSGSWSAGGGSGSTGSTSSSS 699

Qy       3679 TTTTTCTCAAGGACCTTCTTCCAGGCGTGGACGGTCTGCCGGGTGACCCCCAGGTTGAGG 3738
Db        700 SSSSTSTSSSSVSSGSKSSTBSSGSBSSSGSSSSSSTSSBBSCTSTSSSSSSSYSSSTCS 759

Qy       3739 GCCAACATCTCCAGGGGGACCATGAAGACGACCGTCCCCACCTTCCTCCAGAGCTCCCGG 3798
Db        760 CCTCCCSYSYSSSTSSSSSTSWGSTSGSSSSSVGTSSSSDSTSTCCSCCCYMCTCCSTYB 819

Qy       3799 TTGCCGTAGGGGATGGTGGAGCGGGCAATCTCCTGGAGGAGTTCCAGAAGCTTCTGAGGC 3858
Db        820 MBCYTSTSCGGSSSSSGKGGVTKCGCGGCGSSSTNGMBGTSSACSSSSSSCSSSSVSSSS 879

Qy       3859 CCGTCCTGGACGGCTTGACATACATTCCCAACGGGGGGTTCAGCTG 3904
Db        880 KSSASSSSVSSSGSSGVSSNSSSASKSSSSGSVSSGSGSGSGSVSG 925
```

```
RESULT 4
CG352167/c
LOCUS       CG352167                655 bp    DNA     linear   GSS 26-AUG-2003
DEFINITION  OG1BW12TV ZM_0.7_1.5_KB Zea mays genomic clone ZMMBMa0727A23,
            genomic survey sequence.
ACCESSION   CG352167
VERSION     CG352167.1  GI:34269433
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 655)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
            Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
  TITLE     Consortium for Maize Genomics
  JOURNAL   Unpublished (2002)
COMMENT     Other_GSSs: OG1BW12TH
            Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .655
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBMa0727A23"
                     /clone_lib="ZM_0.7_1.5_KB"
                     /note="Vector: pBCSK-; Site_1: HincII; 0.7-1.5 kb
                     methylation filtered genomic DNA library"
ORIGIN

  Query Match             0.9%;  Score 53.2;  DB 29;  Length 655;
  Best Local Similarity   46.1%;  Pred. No. 0.29;
  Matches  178;  Conservative    0;  Mismatches  208;  Indels    0;  Gaps    0;

Qy       3452 CCAGCACGTCCAGGGTGGGGCGGATTCCGTGGTCCTGGTAGGCCTTGACCCAGTTGAAGG 3511
Db        530 CCAGCAGCTGCATGGTGGACGCGGACTCCACGGGGTAGAAGCACGCCACCACGTTGCGGT 471

Qy       3512 AGAGCACGCCGTTGGCCATGTCTAGGGCGAGGTTCCTCCAGGGGTAGATGTAGTCGTCCA 3571
Db        470 CCACCATGGCCACGGCCAGGAACACCGCGAAAGACAGCAGCCCGTGCATGACGTCGCGCA 411

Qy       3572 GGGTGAGCCTGGCTTTCCCTGGCCTCAGCCGGACGGCCCAAAGGGTGCCGATGGCCCGGC 3631
Db        410 CGGTCAGCCTGTACCTGTCGTCCCGCGGCGGCGGCGCGCCGGGGCCGGGGCCGGAGCTGG 351

Qy       3632 GCTCCCCGTTGACGGTTTGGTGAAGGACGTCGGTGGCCACCAGGCCCTTTTTCTCAAGGA 3691
Db        350 AGCCGCCACCCTCGATGGGGAGCAGGCGGCCGATGGGTGTGATGAAGCCGTACCGCAGCG 291

Qy       3692 CCTTCTTCCAGGCGTGGACGGTCTGCCGGGTGACCCCCAGGTTGAGGGCCAACATCTCCA 3751
Db        290 CGCCCGAGGCGGTGTCGCGGTAGCTGTCGGTGAAGCAGAGGACGAAGCAGCAGAGCGCGC 231

Qy       3752 GGGGGACCATGAAGACGACCGTCCCCACCTTCCTCCAGAGCTCCCGGTTGCCGTAGGGGA 3811
Db        230 AGGCGCCGATGAGGCAGCCCGTGAGCGCGCGGTTCGCGGCGTTGCAGGTACCGTACGCGG 171

Qy       3812 TGGTGGAGCGGGCAATCTCCTGGAGG 3837
Db        170 TGAAGGACGGCGACAGCATTTCGAAG 145
```

```
RESULT 5
CC609109
LOCUS       CC609109                679 bp    DNA     linear   GSS 18-JUN-2003
DEFINITION  OGWDJ44TH ZM_0.7_1.5_KB Zea mays genomic clone ZMMBMa0544H16,
            genomic survey sequence.
ACCESSION   CC609109
VERSION     CC609109.1  GI:31970530
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 679)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
            Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
  TITLE     Consortium for Maize Genomics
  JOURNAL   Unpublished (2002)
COMMENT     Other_GSSs: OGWDJ44TV
            Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .679
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBMa0544H16"
                     /clone_lib="ZM_0.7_1.5_KB"
                     /note="Vector: pBCSK-; Site_1: HincII; 0.7-1.5 kb
                     methylation filtered genomic DNA library"
ORIGIN

  Query Match             0.9%;  Score 53.2;  DB 29;  Length 679;
  Best Local Similarity   46.1%;  Pred. No. 0.29;
  Matches  178;  Conservative    0;  Mismatches  208;  Indels    0;  Gaps    0;

Qy       3452 CCAGCACGTCCAGGGTGGGGCGGATTCCGTGGTCCTGGTAGGCCTTGACCCAGTTGAAGG 3511
Db        237 CCAGCAGCTGCCTGGTGGACGCGGACTCCACGGGGTAGAAGCACGCCACCACGTTGCGGT 296

Qy       3512 AGAGCACGCCGTTGGCCATGTCTAGGGCGAGGTTCCTCCAGGGGTAGATGTAGTCGTCCA 3571
Db        297 CCACCATGGCCACGGCCAGGAACACCGCGAACGACAGCAGCCCGTGCATGACGTCGCGCA 356

Qy       3572 GGGTGAGCCTGGCTTTCCCTGGCCTCAGCCGGACGGCCCAAAGGGTGCCGATGGCCCGGC 3631
Db        357 CGGTCAGCCTGTACCTGTCGTCCCGCGGCGGCGGCGCGCCGGGGCCGGGGCCGGAGCTGG 416

Qy       3632 GCTCCCCGTTGACGGTTTGGTGAAGGACGTCGGTGGCCACCAGGCCCTTTTTCTCAAGGA 3691
Db        417 AGCCGCCACCCTCGATGGGGAGCAGGCGGCCGCTGGGTGTGATGAAGCCGTACCGCAGCG 476

Qy       3692 CCTTCTTCCAGGCGTGGACGGTCTGCCGGGTGACCCCCAGGTTGAGGGCCAACATCTCCA 3751
Db        477 CGCCCGTGGCGGTGTCGCGGTAGCTGTCGGTGAAGCAGAGGACGAAGCAGCAGAGCGCGC 536

Qy       3752 GGGGGACCATGAAGACGACCGTCCCCACCTTCCTCCAGAGCTCCCGGTTGCCGTAGGGGA 3811
Db        537 AGGCGCCGATGAGGCAGCCCGTGAGCGCGCGGTTCGCGGCGTTGCAGGTACCGTACGCGG 596

Qy       3812 TGGTGGAGCGGGCAATCTCCTGGAGG 3837
Db        597 TGAAGGACGGCGACAGCATTTCGAAG 622
```

```
RESULT 6
CC661222
LOCUS       CC661222                773 bp    DNA     linear   GSS 19-JUN-2003
DEFINITION  OGLAU18TV ZM_0.7_1.5_KB Zea mays genomic clone ZMMBMa0312C11,
            genomic survey sequence.
ACCESSION   CC661222
VERSION     CC661222.1  GI:32064977
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 773)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
            Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
  TITLE     Consortium for Maize Genomics
  JOURNAL   Unpublished (2002)
COMMENT     Other_GSSs: OGLAU18TH
            Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .773
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBMa0312C11"
                     /clone_lib="ZM_0.7_1.5_KB"
                     /note="Vector: pBCSK-; Site_1: HincII; 0.7-1.5 kb
                     methylation filtered genomic DNA library"
ORIGIN

  Query Match             0.9%;  Score 53.2;  DB 29;  Length 773;
  Best Local Similarity   46.1%;  Pred. No. 0.32;
  Matches  178;  Conservative    0;  Mismatches  208;  Indels    0;  Gaps    0;

Qy       3452 CCAGCACGTCCAGGGTGGGGCGGATTCCGTGGTCCTGGTAGGCCTTGACCCAGTTGAAGG 3511
Db          10 CCAGCAGCTGCCTGGTGGACGCGGACTCCACGGGGTAGAAGCACGCCACCACGTTGCGGT 69

Qy       3512 AGAGCACGCCGTTGGCCATGTCTAGGGCGAGGTTCCTCCAGGGGTAGATGTAGTCGTCCA 3571
Db          70 CCACCATGGCCACGGCCAGGAACACCGCGAACGACAGCAGCCCGTGCATGACGTCGCGCA 129

Qy       3572 GGGTGAGCCTGGCTTTCCCTGGCCTCAGCCGGACGGCCCAAAGGGTGCCGATGGCCCGGC 3631
Db         130 CGGTCAGCCTGTACCTGTCGTCCCGCGGCGGCGGCGCGCCGGGGCCGGGGCCGGAGCTGG 189

Qy       3632 GCTCCCCGTTGACGGTTTGGTGAAGGACGTCGGTGGCCACCAGGCCCTTTTTCTCAAGGA 3691
Db         190 AGCCGCCACCCTCGATGGGGAGCAGGCGGCCGCTGGGTGTGATGAAGCCGTACCGCAGCG 249

Qy       3692 CCTTCTTCCAGGCGTGGACGGTCTGCCGGGTGACCCCCAGGTTGAGGGCCAACATCTCCA 3751
Db         250 CGCCCGTGGCGGTGTCGCGGTAGCTGTCGGTGAAGCAGAGGACGAAGCAGCAGAGCGCGC 309

Qy       3752 GGGGGACCATGAAGACGACCGTCCCCACCTTCCTCCAGAGCTCCCGGTTGCCGTAGGGGA 3811
Db         310 AGGCGCCGATGAGGCAGCCCGTGAGCGCGCGGTTCGCGGCGTTGCAGGTACCGTACGCGG 369

Qy       3812 TGGTGGAGCGGGCAATCTCCTGGAGG 3837
Db         370 TGAAGGACGGCGACAGCATTTCGAAG 395
```

```
RESULT 7
BZ529007/c
LOCUS       BZ529007                807 bp    DNA     linear   GSS 16-DEC-2002
DEFINITION  OGAIR91TC ZM2_0.7_1.5_KB Zea mays genomic clone ZMMBMa0070O14,
            genomic survey sequence.
ACCESSION   BZ529007
VERSION     BZ529007.1  GI:27070951
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 807)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
            Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
  TITLE     Consortium for Maize Genomics
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .807
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBMa0070O14"
                     /clone_lib="ZM2_0.7_1.5_KB"
                     /note="Vector: pBCSK-; Site_1: HincII; 0.7-1.5 kb
                     methylation filtered genomic DNA library"
ORIGIN

  Query Match             0.9%;  Score 53.2;  DB 28;  Length 807;
  Best Local Similarity   46.1%;  Pred. No. 0.32;
  Matches  178;  Conservative    0;  Mismatches  208;  Indels    0;  Gaps    0;

Qy       3452 CCAGCACGTCCAGGGTGGGGCGGATTCCGTGGTCCTGGTAGGCCTTGACCCAGTTGAAGG 3511
             |||||| | | |||||  |  |  | | ||  |  |||  ||||  |
Db        527 CCAGCAGCTGCCTGGTGGACGCGGACTCCACGGGGTAGAAGCACGCCACCACGTTGCGGT 468

Qy       3512 AGAGCACGCCGTTGGCCATGTCTAGGGCGAGGTTCCTCCAGGGGTAGATGTAGTCGTCCA 3571
              | || | |  |||||  |  |  ||||  |  |  ||  |||  ||||  ||
Db        467 CCACCATGGCCACGGCCAGGAACACCGCGAACGACAGCAGCCCGTGCATGACGTCGCGCA 408

Qy       3572 GGGTGAGCCTGGCTTTCCCTGGCCTCAGCCGGACGGCCCAAAGGGTGCCGATGGCCCGGC 3631
              |||  ||||||     |  |  ||  | ||  |     ||  |  ||  |  |  ||  | |
Db        407 CGGTCAGCCTGTACCTGTCGTCCCGCGGCGGCGGCGCGCCGGGGCCGGGGCCGGAGCTGG 348

Qy       3632 GCTCCCCGTTGACGGTTTGGTGAAGGACGTCGGTGGCCACCAGGCCCTTTTTCTCAAGGA 3691
              |  |||     ||  |  ||  | |||  | |||  |||        ||  |     |  ||
Db        347 AGCCGCCACCCTCGATGGGGAGCAGGCGGCCGCTGGGTGTGATGAAGCCGTACCGCAGCG 288

Qy       3692 CCTTCTTCCAGGCGTGGACGGTCTGCCGGGTGACCCCCAGGTTGAGGGCCAACATCTCCA 3751
              |    | |   ||  ||  |  |        |||||  |  |||  ||  |     |  | |
Db        287 CGCCCGTGGCGGTGTCGCGGTAGCTGTCGGTGAAGCAGAGGACGAAGCAGCAGAGCGCGC 228

Qy       3752 GGGGGACCATGAAGACGACCGTCCCCACCTTCCTCCAGAGCTCCCGGTTGCCGTAGGGGA 3811
              ||  |  ||||  |  |  ||||    | |     ||  |  |  |  |  |||||  | |
Db        227 AGGCGCCGATGAGGCAGCCCGTGAGCGCGCGGTTCGCGGCGTTGCAGGTACCGTACGCGG 168

Qy       3812 TGGTGGAGCGGGCAATCTCCTGGAGG 3837
              ||  |||  |  |  |  |  | ||||
Db        167 TGAAGGACGGCGACAGCATTTCGAAG 142
```

```
RESULT 8
CG034209/c
LOCUS       CG034209                815 bp    DNA     linear   GSS 19-AUG-2003
DEFINITION  PUJGY50TD ZM_0.6_1.0_KB Zea mays genomic clone ZMMBTa0666I03,
            genomic survey sequence.
ACCESSION   CG034209
VERSION     CG034209.1  GI:33906365
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 815)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
            Bennetzen,J.
  TITLE     Maize Genomics Consortium
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .815
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBTa0666I03"
                     /clone_lib="ZM_0.6_1.0_KB"
                     /note="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
                     CoT selected genomic DNA library"
ORIGIN

  Query Match             0.9%;  Score 53.2;  DB 29;  Length 815;
  Best Local Similarity   46.1%;  Pred. No. 0.33;
  Matches  178;  Conservative    0;  Mismatches  208;  Indels    0;  Gaps    0;

Qy       3452 CCAGCACGTCCAGGGTGGGGCGGATTCCGTGGTCCTGGTAGGCCTTGACCCAGTTGAAGG 3511
             |||||| | | |||||  |  |  | | ||  |  |||  ||||  |
Db        523 CCAGCAGCTGCCTGGTGGACGCGGACTCCACGGGGTAGAAGCACGCCACCACGTTGCGGT 464

Qy       3512 AGAGCACGCCGTTGGCCATGTCTAGGGCGAGGTTCCTCCAGGGGTAGATGTAGTCGTCCA 3571
              | || | |  |||||  |  |  ||||  |  |  ||  |||  ||||  ||
Db        463 CCACCATGGCCACGGCCAGGAACACCGCGAACGACAGCAGCCCGTGCATGACGTCGCGCA 404

Qy       3572 GGGTGAGCCTGGCTTTCCCTGGCCTCAGCCGGACGGCCCAAAGGGTGCCGATGGCCCGGC 3631
              |||  ||||||     |  |  ||  | ||  |     ||  |  ||  |  |  ||  | |
Db        403 CGGTCAGCCTGTACCTGTCGTCCCGCGGCGGCGGCGCGCCGGGGCCGGGGCCGGAGCTGG 344

Qy       3632 GCTCCCCGTTGACGGTTTGGTGAAGGACGTCGGTGGCCACCAGGCCCTTTTTCTCAAGGA 3691
              |  |||     ||  |  ||  | |||  | |||  |||        ||  |     |  ||
Db        343 AGCCGCCACCCTCGATGGGGAGCAGGCGGCCGCTGGGTGTGATGAAGCCGTACCGCAGCG 284

Qy       3692 CCTTCTTCCAGGCGTGGACGGTCTGCCGGGTGACCCCCAGGTTGAGGGCCAACATCTCCA 3751
              |    | |   ||  ||  |  |        |||||  |  |||  ||  |     |  | |
Db        283 CGCCCGTGGCGGTGTCGCGGTAGCTGTCGGTGAAGCAGAGGACGAAGCAGCAGAGCGCGC 224

Qy       3752 GGGGGACCATGAAGACGACCGTCCCCACCTTCCTCCAGAGCTCCCGGTTGCCGTAGGGGA 3811
              ||  |  ||||  |  |  ||||    | |     ||  |  |  |  |  |||||  | |
Db        223 AGGCGCCGATGAGGCAGCCCGTGAGCGCGCGGTTCGCGGCGTTGCAGGTACCGTACGCGG 164

Qy       3812 TGGTGGAGCGGGCAATCTCCTGGAGG 3837
              ||  |||  |  |  |  |  | ||||
Db        163 TGAAGGACGGCGACAGCATTTCGAAG 138

RESULT 9
```

```
CC661216/c
LOCUS       CC661216                897 bp    DNA     linear   GSS 19-JUN-2003
DEFINITION  OGLAU18TH ZM_0.7_1.5_KB Zea mays genomic clone ZMMBMa0312C11,
            genomic survey sequence.
ACCESSION   CC661216
VERSION     CC661216.1  GI:32064967
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 897)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
            Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
  TITLE     Consortium for Maize Genomics
  JOURNAL   Unpublished (2002)
COMMENT     Other_GSSs: OGLAU18TV
            Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .897
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBMa0312C11"
                     /clone_lib="ZM_0.7_1.5_KB"
                     /note="Vector: pBCSK-; Site_1: HincII; 0.7-1.5 kb
                     methylation filtered genomic DNA library"
ORIGIN

  Query Match             0.9%;  Score 53.2;  DB 29;  Length 897;
  Best Local Similarity   46.1%;  Pred. No. 0.35;
  Matches  178;  Conservative    0;  Mismatches  208;  Indels    0;  Gaps    0;

Qy       3452 CCAGCACGTCCAGGGTGGGGCGGATTCCGTGGTCCTGGTAGGCCTTGACCCAGTTGAAGG 3511
Db        797 CCAGCAGCTGCCTGGTGGACGCGGACTCCACGGGGTAGAAGCACGCCACCACGTTGCGGT 738

Qy       3512 AGAGCACGCCGTTGGCCATGTCTAGGGCGAGGTTCCTCCAGGGGTAGATGTAGTCGTCCA 3571
Db        737 CCACCATGGCCACGGCCAGGAACACCGCGAACGACAGCAGCCCGTGCATGACGTCGCGCA 678

Qy       3572 GGGTGAGCCTGGCTTTCCCTGGCCTCAGCCGGACGGCCCAAAGGGTGCCGATGGCCCGGC 3631
Db        677 CGGTCAGCCTGTACCTGTCGTCCCGCGGCGGCGGCGCGCCGGGGCCGGGGCCGGAGCTGG 618

Qy       3632 GCTCCCCGTTGACGGTTTGGTGAAGGACGTCGGTGGCCACCAGGCCCTTTTTCTCAAGGA 3691
Db        617 AGCCGCCACCCTCGATGGGGAGCAGGCGGCCGCTGGGTGTGATGAAGCCGTACCGCAGCG 558

Qy       3692 CCTTCTTCCAGGCGTGGACGGTCTGCCGGGTGACCCCCAGGTTGAGGGCCAACATCTCCA 3751
Db        557 CGCCCGTGGCGGTGTCGCGGTAGCTGTCGGTGAAGCAGAGGACGAAGCAGCAGAGCGCGC 498

Qy       3752 GGGGGACCATGAAGACGACCGTCCCCACCTTCCTCCAGAGCTCCCGGTTGCCGTAGGGGA 3811
Db        497 AGGCGCCGATGAGGCAGCCCGTGAGCGCGCGGTTCGCGGCGTTGCAGGTACCGTACGCGG 438

Qy       3812 TGGTGGAGCGGGCAATCTCCTGGAGG 3837
Db        437 TGAAGGACGGCGACAGCATTTCGAAG 412

RESULT 10
```

```
BX381961
LOCUS       BX381961                1201 bp    mRNA    linear   EST 08-MAY-2003
DEFINITION  BX381961 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI072YF05 3-PRIME, mRNA sequence.
ACCESSION   BX381961
VERSION     BX381961.1  GI:30453007
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DI072CC03NP1.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI072YF05"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

  Query Match             0.9%;  Score 52.4;  DB 13;  Length 1201;
  Best Local Similarity   5.7%;  Pred. No. 0.64;
  Matches   44;  Conservative  279;  Mismatches  452;  Indels    1;  Gaps    1;

Qy       2906 ACTCCAACCAAGGAATCCGCCAAAGGCGGCCTACCTTTTGAGCCCGTATCTTCCCCTGAC 2965
Db        399 ANNNNNKCKMMMNNNKKKKKCMNKNKKKMMNKKNMNKKKKKKKKNNNNNCNANNMMNNKN 458

Qy       2966 GTATAGACCTTCGGATCGTCTCAGGGTGCACCCGAAGGATGTCTGCAAGCTCCTCGGGGG 3025
Db        459 KKKNNKNNNNTKNNMNNNNCNKMNNKNKKNNNNNKMNKNMMNKNNNCNNNKMCKMMNMMN 518

Qy       3026 TCAGGTACACGGGCTTCATCCTCATGACACAACCTTACCCCACAGAGGACAACACATGCA 3085
Db        519 KMMMNKMMNNNCNKMMNAMNNKKMNKMMMNKNNMMNCKTNNNNNNNNKNNNNNKNNNNN 578

Qy       3086 ACTATGGGCAAAGTAGACAACGAGACCAAAAGCTTGGGCCACTCTCTCAGGAGGCCTCCT 3145
Db        579 KKKKKKKNNNNNKNNNMNNNNANKNNNNNNNNNKNNCMNKKKKKKMKMMMMNKMNNNNN 638

Qy       3146 TGAGGGTCTTCACTAGGTACGCTCCCGGTCGTGTCAGATGGCCATCCGTGTAATCTCGGA 3205
Db        639 KKNKNNNKNNNNMNKMNNNNNKNNNMNNNNNNNNNNNNNNKKKKKKKKNNNNNNNNNNN 698

Qy       3206 TAACCCGCATTAGGACGGCAAATAGATATTGCGCGGGGAGTTCACCCCTGGCCACAGCCC 3265
Db        699 NMKMNNNNKKNKKTNKKKKNNAKKNNTNNKMMNKNNNNNNMMNNCNCNKKKNNKTNMNN 758

Qy       3266 ACAGCAAGCCTGCATAGAAACGTCTTGAACGACGGTCATCTAGGAGATCGGCAATGTACG 3325
Db        759 MMNNNNNKKNNNNNKNNNNNNNNNNKKMKMSKCKKKKMCCKMCCMCKKKKKMBKGKMVCM 818

Qy       3326 TAGCAATGAGGGTGATAAGGGCCGGGAGTTTGGAACGCTCCACCTCGGGGAGGACCAGGA 3385
Db        819 CKMMMKNKKCMCMKBMMMCKMCMKMCMBKKNMMMMMMKMKMMMNKNKAMMKKDNMMMMM 878
```

```
Qy       3386 TGAGGCCCAGGTCAACGGCCACGGTCTTGGTGTTGGGCATCACCCTTTTCCCCTGAGCCC 3445
Db        879 MCAMMCMMMKKNMMMBKKNCMMKKMCCMMKKKKHKKNMMKKMMMCKKBMMCKKNKMMNKM 938

Qy       3446 AGAGGACCAGCACGTCCAGGGTGGGGCGGATTCCGTGGTCCTGGTAGGCCTTGACCCAGT 3505
Db        939 MVKKKKAMMNBKCV-MKMKMKMCCKKVMMBKMCAKKKKMCMCVKKVMCMCDBKMCMCKKK 997

Qy       3506 TGAAGGAGAGCACGCCGTTGGCCATGTCTAGGGCGAGGTTCCTCCAGGGGTAGATGTAGT 3565
Db        998 BMKKVCCKKMCMMMMMCAMKVMMMBAMMMMBAMVVMMMMNNMMMMMCKKMVMKKKKMCK 1057

Qy       3566 CGTCCAGGGTGAGCCTGGCTTTCCCTGGCCTCAGCCGGACGGCCCAAAGGGTGCCGATGG 3625
Db       1058 KKBMSKKMMMNKKVCKMBBMCMSKGGCMMCGCMCKGCVGCMNSCCMMBGSKGCCKKKMBK 1117

Qy       3626 CCCGGCGCTCCCCGTTGACGGTTTGGTGAAGGACGTCGGTGGCCACCAGGCCCTTT 3681
Db       1118 CMMCCMCMCBCMCBKKVCVKKKKKCKKKKKKKKCKKCCNMKKBCKKCKKKKBKKKK 1173
```

RESULT 11
CNS0091P/c

```
LOCUS       CNS0091P                925 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL053013
VERSION     AL053013.1  GI:4934461
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 925)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .925
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR19D16"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match             0.9%;  Score 51.8;  DB 29;  Length 925;
  Best Local Similarity   13.7%;  Pred. No. 0.77;
  Matches   54;  Conservative  172;  Mismatches  167;  Indels    0;  Gaps    0;
```

```
Qy       3363 CTCCACCTCGGGGAGGACCAGGATGAGGCCCAGGTCAACGGCCACGGTCTTGGTGTTGGG 3422
Db        924 SBSCSCSCSCSSBSCSSSSMSTSSSNSSBCSSCSSSBSSSSTSSMSSSSBSSSSGSSSSS 865

Qy       3423 CATCACCCTTTTCCCCTGAGCCCAGAGGACCAGCACGTCCAGGGTGGGGCGGATTCCGTG 3482
Db        864 SGTSSACVKCNASSSCGCCGCGMABCCMCSSSSSCCGSASARGVKVRASGGAGKRGGGSG 805

Qy       3483 GTCCTGGTAGGCCTTGACCCAGTTGAAGGAGAGCACGCCGTTGGCCATGTCTAGGGCGAG 3542
Db        804 GASASHSSSSACBSSSSSCSASCWSASSSSSASSSRSRSGGGAGGSGASSSRSSSSSSSA 745

Qy       3543 GTTCCTCCAGGGGTAGATGTAGTCGTCCAGGGTGAGCCTGGCTTTCCCTGGCCTCAGCCG 3602
Db        744 SAGSVVSSASSSSSSCSSSVSCSSVASSMSCSSBSSSSASASSSSSSSSSASCASCSCCCT 685

Qy       3603 GACGGCCCAAAGGGTGCCGATGGCCCGGCGCTCCCCGTTGACGGTTTGGTGAAGGACGTC 3662
Db        684 SWSCSCSTSASMSAARSSSSSSSSCSSSSMSASSSASSSSASSSSSSGSSSSSGSACGBS 625

Qy       3663 GGTGGCCACCAGGCCCTTTTTCTCAAGGACCTTCTTCCAGGCGTGGACGGTCTGCCGGGT 3722
Db        624 MSSGGGSGSVSASSGMSSSVSSSGGRSSGSGGGGGVGGSGSSSSGSGSSGGSGSVCSCSS 565

Qy       3723 GACCCCCAGGTTGAGGGCCAACATCTCCAGGGG 3755
Db        564 GCMCRCSCSSAAAAAASCVAASCGMMCGKSKSG 532
```

RESULT 12
CA552474/c

```
LOCUS       CA552474                497 bp    mRNA    linear   EST 19-NOV-2002
DEFINITION  C0857F07-5N NIA Mouse Blastocyst cDNA Library (Long) Mus musculus
            cDNA clone NIA:C0857F07 IMAGE:30030498 5', mRNA sequence.
ACCESSION   CA552474
VERSION     CA552474.1  GI:25096660
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 497)
  AUTHORS   Piao,Y., Kargul,G.J., Dudekula,D.B., Qian,Y., Tanaka,T., Luo,A. and
            Ko,M.S.H.
  TITLE     Systematic Analyses of NIA Mouse Blastocyst cDNA Library (Long)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Dawood B. Dudekula
            Laboratory of Genetics
            National Institute on Aging/National Institutes of Health
            333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
            Email: cdna@lgsun.grc.nia.nih.gov
            Plate: C0857  row: F  column: 07
            Seq primer: M13 Reverse
            High quality sequence stop: 497
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .497
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="niaEST:C0857F07-5N"
                     /db_xref="taxon:10090"
                     /clone="NIA:C0857F07 IMAGE:30030498"
                     /tissue_type="Blastocyst"
                     /dev_stage="3.5-dpc"
                     /lab_host="DH10B"
                     /clone_lib="NIA Mouse Blastocyst cDNA Library (Long)"
                     /note="Vector: pSPORT1 (Invitrogen); Site_1: SalI; Site_2:
                     NotI; Mouse cDNA project by the Laboratory of Genetics,
                     National Institute on Aging (NIA), Intramural Research
                     Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA). This is
                     a long-transcript enriched cDNA library (Ref. Genome Res.
                     11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were
                     extracted from a pool of 20 Blastocysts. Double-stranded
                     cDNAs were synthesized with an Oligo(dT) primer
                     [Invitrogen:
```

```
                  5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCCTTTTTTTTTTTTTT-3'] from
                  0.2 ug of total RNA, treated with T4 DNA polymerase, and
                  purified by ethanol-precipitation. The cDNAs were ligated
                  to Lone-linker LL-Sal4, purified by phenol/chloroform, and
                  separated from free linkers by Centricon 100. Then, the
                  cDNAs were amplified by long-range high fidelity PCR using
                  Ex Taq polymerase (Takara) with a primer Sal4-S. The
                  products were purified by phenol/chloroform and Centricon
                  100. The cDNAs were digested with SalI and NotI enzymes
                  and cloned into SalI/NotI site of pSPORT1 plasmid vector.
                  The DH10B E. coli host was transformed with the ligation
                  mixture by the standard chemical method. The average
                  insert size is about 2.2 kb. The library was constructed
                  by Yulan Piao (NIA)."
ORIGIN

  Query Match             0.9%;  Score 51.6;  DB 14;  Length 497;
  Best Local Similarity   56.5%;  Pred. No. 0.6;
  Matches   96;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy       2288 ACGAAGTTTCGCCGGCCAAAGACCTCATCAAGGATGAGCTTCACCTCGAACCCGTATTTC 2347
               | ||| || ||||| ||||| |||||||| |  |  |  || |  |       || |  ||
Db        496 ATGAAATTCCGCCGACCAAACACCTCATCCATAACAACTTTGAGGTAATGTCCCTCATTG 437

Qy       2348 TCGTCTATGTGAACGAAGATCAGTCCTGAGTCCGCCATCAGCTCCCTGAGAAGTATCAAG 2407
              ||||| ||      || ||||     |||     ||||||| |||||| |   ||     |
Db        436 TCGTCAATCGACACCCAGATGCTGCCTTCCTCCGCCAACAGCTCACGTAACAGGGCGAGA 377

Qy       2408 CGCTCCCTCAGGAACTCCACAAACTGAGGACCATCGAGGGTGTCATCGTA 2457
               |||      |  | | | |||  ||||        || ||  |||| |||||
Db        376 CGCGGATAGATCATCGCCAGCCACTGGCTGTGTTCCAGATTGTCGTCGTA 327


RESULT 13
CA230834/c
LOCUS       CA230834                 649 bp    mRNA    linear   EST 25-SEP-2003
DEFINITION  SCJFFL3C07G11.g Saccharum officinarum FL3 Saccharum officinarum
            cDNA clone SCJFFL3C07G11 5', mRNA sequence.
ACCESSION   CA230834
VERSION     CA230834.1  GI:35294765
KEYWORDS    EST.
SOURCE      Saccharum officinarum
  ORGANISM  Saccharum officinarum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Saccharum.
REFERENCE   1  (bases 1 to 649)
  AUTHORS   Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.
  TITLE     The libraries that made SUCEST
  JOURNAL   Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
COMMENT     Contact: Arruda P
            Centro de Biologia Molecular e Engenharia Genetica
            Universidade Estadual de Campinas
            Caixa Postal 6010, 13083-970, Campinas SP, Brazil
            Tel: 55 19 3788 1137
            Fax: 55 19 3788 1089
            Email: parruda@unicamp.br
            Clone distribution: clone distribution information can be found
            through the Brazilian Clone Collection Center (BCCC) at
            http://www.bcccenter.fcav.unesp.br
            Plate: C07  row: G  column: 11
            Seq primer: T7 Promoter Primer.
FEATURES             Location/Qualifiers
     source          1..649
                     /organism="Saccharum officinarum"
                     /mol_type="mRNA"
                     /db_xref="taxon:4547"
                     /clone="SCJFFL3C07G11"
                     /lab_host="DH10B"
                     /clone_lib="Saccharum officinarum FL3"
                     /note="Organ: Base of developing inflorescence (5cm-long);
                     Vector: pSport1; Site_1: SalI; Site_2: NotI; An
                     unidirectional cDNA library generated from [Base of
                     developing  inflorescence (5cm-long)]. cDNA was prepared
                     from polyA+ mRNA using  SuperScript Plasmid System Kit
                     (Invitrogen). The double-strand cDNAs were  fractionated
                     in a sepharose CL-2B 40cm-columns and fragments sizing
                     between 0.8 and 1.5 Kb were directionally cloned into the
                     vector. Details  of each source of RNA and library
                     construction can be obtained at
                     http://sucest.lad.ic.unicamp.br/public"
ORIGIN

  Query Match             0.9%;  Score 51.2;  DB 14;  Length 649;
  Best Local Similarity   43.9%;  Pred. No. 0.87;
  Matches  218;  Conservative    0;  Mismatches  279;  Indels    0;  Gaps    0;

Qy       3333 GAGGGTGATAAGGGCCGGGAGTTTGGAACGCTCCACCTCGGGGAGGACCAGGATGAGGCC 3392
              ||||  |   |  | |   |  | ||  |   | |    ||||  |    |||||    |
Db        610 GAGGCGGGCAGTGACGTTGNATATGCAGATCCCGGTGTAGGGGTCGTTGGGGATGCCCTC 551

Qy       3393 CAGGTCAACGGCCACGGTCTTGGTGTTGGGCATCACCCTTTTCCCCTGAGCCCAGAGGAC 3452
              ||      |||||| ||||  |  |  |  |  |  |||       |     |  | |||
Db        550 CATCCTGCCGGCCATGGTCACGTTCTCAGCGAACACGTCGCTGTAGTTGATGCCGGTGAC 491

Qy       3453 CAGCACGTCCAGGGTGGGGCGGATTCCGTGGTCCTGGTAGGCCTTGACCCAGTTGAAGGA 3512
               |          | |||||  |      || ||| |||      |     |    |
Db        490 CTCGGGCAAGGCGTTGGGGTTGGACGTGTTGTCGGGGTGCTGCCCGTAGTTGCCGGTCAT 431

Qy       3513 GAGCACGCCGTTGGCCATGTCTAGGGCGAGGTTCCTCCAGGGGTAGATGTAGTCGTCCAG 3572
                |   |      |||   ||   |   |    |   |    | | ||||    | | |
Db        430 CCAGAACACCCACTTCATGGTGTGGAGGCTGAGGCCGCGCACGAAAATGTCCCTGACAAA 371

Qy       3573 GGTGAGCCTGGCTTTCCCTGGCCTCAGCCGGACGGCCCAAAGGGTGCCGATGGCCCGGCG 3632
               |  |  ||  |      || |  |  | | || |||||||      |||  ||||||  |
Db        370 GCCGCCCCTCCCCGCGCCGGACTTGACCCTGACGGCCGACTCCGTGTTGATGGCGACGTT 311

Qy       3633 CTCCCCGTTGACGGTTTGGTGAAGGACGTCGGTGGCCACCAGGCCCTTTTTCTCAAGGAC 3692
               ||| ||  |        | |  |  |  | || |||     |    ||||        | | |
Db        310 GTCCTCGGCGCGCACGTCGCGGATGCCGCCGGACATCTCGCTGCCCAGCGCGATCATGGC 251

Qy       3693 CTTCTTCCAGGCGTGGACGGTCTGCCGGGTGACCCCCAGGTTGAGGGCCAACATCTCCAG 3752
                || |    || |  |  |||| |||    |||    | | ||    ||  |  ||| |  |
Db        250 GCTCGTGGGGGAGATGCAGGTCAGCCTCCTGATGACGATGTGCTGGCTCGGCATGTTGAA 191

Qy       3753 GGGGACCATGAAGACGACCGTCCCCACCTTCCTCCAGAGCTCCCGGTTGCCGTAGGGGAT 3812
                 ||   |  |  ||  ||   |||   ||||      ||           || ||||  ||  |||
Db        190 CCTGATGCCGTACTCGTCCCACCCGCTCTTCACGGCGACGCAGTCGTCGCCGGAGACGAT 131

Qy       3813 GGTGGAGCGGGCAATCT 3829
              |  | ||    | ||||
Db        130 GTAGCAGTCCTCGATCT 114


RESULT 14
CB870888/c
LOCUS       CB870888                 649 bp    mRNA    linear   EST 03-JUL-2003
DEFINITION  HC15J04w CH Hordeum vulgare cDNA clone HC15J04 5-PRIME, mRNA
            sequence.
ACCESSION   CB870888
VERSION     CB870888.1  GI:30072868
KEYWORDS    EST.
SOURCE      Hordeum vulgare
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 649)
  AUTHORS   Varshney,R.K., Zhang,H., Burton,R., Stein,N., Langridge,P. and
            Graner,A.
  TITLE     Barley ESTs from coleoptile tissue
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Stein Nils
            Molecular Markers Group, Department Genbank
            Institute of Plant Genetics and Crop Plant Research (IPK)
            Corrensstr. 3, 06466, Gatersleben, Germany
```

Tel: 039482-5522
Fax: 039482-5595
Email: stein@ipk-gatersleben.de
Insert Length: 649 Std Error: 0.00
Plate: 15 row: J column: 4
Seq primer: T7.

FEATURES Location/Qualifiers
source 1. .649
/organism="Hordeum vulgare"
/mol_type="mRNA"
/cultivar="Sloop"
/db_xref="GABI:549289"
/db_xref="taxon:4513"
/clone="HC15J04"
/tissue_type="coleoptile"
/dev_stage="coleoptile, 3 days old"
/lab_host="DH10B"
/clone_lib="CH"
/note="Vector: pSPORT; Site_1: SalI (5-end of cDNA); Site_2: NotI (3-end of cDNA); Due to the cloning system used blue/white selection for recombinants is not 100% reliable. Average insert size is 1.3 kb."

ORIGIN

Query Match 0.9%; Score 51.2; DB 14; Length 649;
Best Local Similarity 46.3%; Pred. No. 0.87;
Matches 167; Conservative 0; Mismatches 194; Indels 0; Gaps 0;

```
Qy       3351 GAGTTTGGAACGCTCCACCTCGGGGAGGACCAGGATGAGGCCCAGGTCAACGGCCACGGT 3410
             ||  | |   ||||   |  ||| |||| ||  | ||||| | ||| | | | |
Db        587 GATCTCGACGCGCTGGTCGNCGGCGAGGGCCTTGTTGAGGTCGAGGACGAAGTCGCGGTA 528

Qy       3411 CTTGGTGTTGGGCATCACCCTTTTCCCCTGAGCCCAGAGGACCAGCACGTCCAGGGTGGG 3470
               | |  | ||  | |   |  |  ||  ||| || |    | |      | |
Db        527 GTAGCGGATGTACTTGCGCATGGGGGCGTCGGCGGGGAGCACGACGGAGCCGTTCCAGAG 468

Qy       3471 GCGGATTCCGTGGTCCTGGTAGGCCTTGACCCAGTTGAAGGAGAGCACGCCGTTGGCCAT 3530
             |  | |  ||| | |  ||  | |   ||||  |  ||      | | ||  | |
Db        467 GGTGTTGTCGTAGCCGAGGAGGCCGCCGACCTTGACGAGCTTCATGAGGCGCTCGTGGTA 408

Qy       3531 GTCTAGGGCGAGGTTCCTCCAGGGGTAGATGTAGTCGTCCAGGGTGAGCCTGGCTTTCCC 3590
             ||  |||  |  || | |  || ||  | | || ||      |  | |  ||  |
Db        407 GTTGAGGTAGTTGTCCTTGTCGGCGTCCACGAAGACGAAGTCGAAGGTGCCGTGGTTGGC 348

Qy       3591 TGGCCTCAGCCGGACGGCCCAAAGGGTGCCGATGGCCCGGCGCTCCCCGTTGACGGTTTG 3650
                  |||  |||   |||  |  ||  ||| ||| ||| | |  | ||| | |
Db        347 CTCGTCCTCGAGGAGGGCGTCCAGGACGGGGAGCGCCGGGCCCTCGCGGAAGTCGATCTT 288

Qy       3651 GTGAAGGACGTCGGTGGCCACCAGGCCCTTTTTCTCAAGGACCTTCTTCCAGGCGTGGAC 3710
             |||    ||| |||    | | | ||  |    | | ||  | |  |||  |  || ||
Db        287 GTGCGCCACGCCGGCCTTCTCGATGCACGGCAGCCCCAGCTCGTAGTTCTCGCGGTTGAT 228

Qy       3711 G 3711
              |
Db        227 G 227
```

RESULT 15
CB874255
LOCUS CB874255 594 bp mRNA linear EST 03-JUL-2003
DEFINITION HC15J04y CH Hordeum vulgare cDNA clone HC15J04 3-PRIME, mRNA sequence.
ACCESSION CB874255
VERSION CB874255.1 GI:30076241
KEYWORDS EST.
SOURCE Hordeum vulgare
ORGANISM Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 594)
AUTHORS Varshney,R.K., Zhang,H., Burton,R., Stein,N., Langridge,P. and Graner,A.
TITLE Barley ESTs from coleoptile tissue
JOURNAL Unpublished (2003)
COMMENT Contact: Stein Nils
Molecular Markers Group, Department Genbank
Institute of Plant Genetics and Crop Plant Research (IPK)
Correnssstr. 3, 06466, Gatersleben, Germany
Tel: 039482-5522
Fax: 039482-5595
Email: stein@ipk-gatersleben.de
Insert Length: 594 Std Error: 0.00
Plate: 15 row: J column: 4
Seq primer: SP6.

FEATURES Location/Qualifiers
source 1. .594
/organism="Hordeum vulgare"
/mol_type="mRNA"
/cultivar="Sloop"
/db_xref="GABI:552657"
/db_xref="taxon:4513"
/clone="HC15J04"
/tissue_type="coleoptile"
/dev_stage="coleoptile, 3 days old"
/lab_host="DH10B"
/clone_lib="CH"
/note="Vector: pSPORT; Site_1: SalI (5-end of cDNA); Site_2: NotI (3-end of cDNA); Due to the cloning system used blue/white selection for recombinants is not 100% reliable. Average insert size is 1.3 kb."

ORIGIN

Query Match 0.9%; Score 50.6; DB 14; Length 594;
Best Local Similarity 46.3%; Pred. No. 1.2;
Matches 167; Conservative 0; Mismatches 194; Indels 0; Gaps 0;

```
Qy       3351 GAGTTTGGAACGCTCCACCTCGGGGAGGACCAGGATGAGGCCCAGGTCAACGGCCACGGT 3410
             ||  | |   ||||   |  ||| |||| ||  | ||||| | ||| | | | |
Db        228 GATCTCGACGCGCTGGTCGGCGGCGAGGGCCTTGTTGAGGTCGAGGACGAAGTCGCGGTA 287

Qy       3411 CTTGGTGTTGGGCATCACCCTTTTCCCCTGAGCCCAGAGGACCAGCACGTCCAGGGTGGG 3470
               | |  | ||  | |   |  |  ||  ||| || |    | |      | |
Db        288 GTAGCGGATGTACTTGCGCATGGGGGCGTCGGCGGGGAGCACGACGGAGCCGTTCCAGAG 347

Qy       3471 GCGGATTCCGTGGTCCTGGTAGGCCTTGACCCAGTTGAAGGAGAGCACGCCGTTGGCCAT 3530
             |  | |  ||| | |  ||  | |   ||||  |  ||      | | ||  | |
Db        348 GGTGTTGTCGTAGCCGAGGAGGCCGCCGACCTTGACGAGCTTCATGAGGCGCTCGTGGTA 407

Qy       3531 GTCTAGGGCGAGGTTCCTCCAGGGGTAGATGTAGTCGTCCAGGGTGAGCCTGGCTTTCCC 3590
             ||  |||  |  || | |  || ||  | | || ||      |  | |  ||  |
Db        408 GTTGAGGTAGTTGTCCTTGTCGGCGTCCACGAAGACGAAGTCGAAGGTGCCGTGGTTGGC 467

Qy       3591 TGGCCTCAGCCGGACGGCCCAAAGGGTGCCGATGGCCCGGCGCTCCCCGTTGACGGTTTG 3650
                  |||  |||   |||  |  ||  ||| ||| ||| | |  | ||| | |
Db        468 CTCGTCCTCGAGGAGGGCGTCCAGGACGGGGAGCGCCGGGCCCTCGCGGAAGTCGATCTT 527

Qy       3651 GTGAAGGACGTCGGTGGCCACCAGGCCCTTTTTCTCAAGGACCTTCTTCCAGGCGTGGAC 3710
             |||    ||| |||    | | | ||  |    | | ||  | |  |||  |  || ||
Db        528 GTGCGCCACGCCGGCCTTCTCGATGCACGGCAGCCCCAGCTCGTAGTTCTCGCGGTTGAT 587

Qy       3711 G 3711
              |
Db        588 G 588
```

Search completed: March 18, 2004, 05:45:37
Job time : 9469.64 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: March 17, 2004, 20:21:47 ; Search time 2608.35 Seconds
(without alignments)
17049.054 Million cell updates/sec

Title: US-09-664-186-4
Perfect score: 1026
Sequence: 1 gtgaagaacgaaaaaacctt.....ccctcaaggaggcctcctga 1026

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1026	100.0	1026	6	AR139424	AR139424 Sequence
2	1026	100.0	1600	1	U89376	U89376 Thermus sp.
c 3	1026	100.0	5849	1	AF013571	AF013571 Thermus s
c 4	1026	100.0	5849	6	AR139426	AR139426 Sequence
5	59.8	5.8	308050	1	SCO939124	AL939124 Streptomy
c 6	56.8	5.5	2000	6	AX655393	AX655393 Sequence
c 7	56.8	5.5	300800	1	SCO939112	AL939112 Streptomy
8	56.4	5.5	2030	6	AX653795	AX653795 Sequence
9	56.4	5.5	2031	6	AX755741	AX755741 Sequence
10	55.4	5.4	2028	6	AX660328	AX660328 Sequence
11	55.4	5.4	2028	6	AX755777	AX755777 Sequence
12	54.8	5.3	2028	6	AX653794	AX653794 Sequence
13	54.8	5.3	2060	6	AX654685	AX654685 Sequence
14	54.8	5.3	2273	8	AK106696	AK106696 Oryza sat
c 15	54.8	5.3	137651	8	AC130600	AC130600 Oryza sat
c 16	54.8	5.3	177374	8	AC135429	AC135429 Oryza sat
17	54.6	5.3	299050	1	SCO939119	AL939119 Streptomy
c 18	53.8	5.2	299425	1	AP005037	AP005037 Streptomy
19	53.6	5.2	303550	1	SCO939118	AL939118 Streptomy
20	53.2	5.2	1755	6	AX654077	AX654077 Sequence
21	53.2	5.2	1965	6	AX659936	AX659936 Sequence
22	53.2	5.2	1965	6	AX755739	AX755739 Sequence
23	53.2	5.2	2303	8	AK072823	AK072823 Oryza sat
24	53.2	5.2	100800	8	AC105260	AC105260 Oryza sat
25	53.2	5.2	142854	8	AC136216	AC136216 Oryza sat
c 26	53.2	5.2	150188	8	AP005499	AP005499 Oryza sat
c 27	53.2	5.2	167446	8	AP004746	AP004746 Oryza sat
c 28	52.6	5.1	298450	1	SCO939107	AL939107 Streptomy
29	52.4	5.1	113193	1	AF357202	AF357202 Streptomy
30	52.4	5.1	113193	6	AX703543	AX703543 Sequence
c 31	52.4	5.1	321250	1	SCO939111	AL939111 Streptomy
32	52.2	5.1	3297	6	BD180129	BD180129 Highly th
33	52.2	5.1	188050	1	AL646072	AL646072 Ralstonia
34	51.8	5.0	19372	1	AE004443	AE004443 Pseudomon
35	51.6	5.0	1815	6	AX148314	AX148314 Sequence
36	51.6	5.0	1815	6	BD017435	BD017435 Transgeni
37	51.6	5.0	2498	8	ZMU95953	U95953 Zea mays vi
c 38	51.6	5.0	138203	1	AY310323	AY310323 Streptomy
39	51.2	5.0	615	6	E04306	E04306 DNA encodin
40	51.2	5.0	915	1	TTHSUPDIS	D13387 Thermus aqu
41	51.2	5.0	6955	1	TTHILVB	D84646 Thermus aqu
42	50.8	5.0	2283	6	BD180126	BD180126 Highly th
43	50.8	5.0	276800	1	SCO939115	AL939115 Streptomy
44	50.8	5.0	298550	1	AP005047	AP005047 Streptomy
c 45	50.8	5.0	301617	1	AE016911	AE016911 Chromobac

ALIGNMENTS

RESULT 1
AR139424
LOCUS AR139424 1026 bp DNA linear PAT 16-JUN-2001
DEFINITION Sequence 4 from patent US 6207377.
ACCESSION AR139424
VERSION AR139424.1 GI:14481920
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 1026)
AUTHORS Wayne,J. and Xu,S.-y.
TITLE Method for construction of thermus-E. coli shuttle vectors and identification of two Thermus plasmid replication origins
JOURNAL Patent: US 6207377-A 4 27-MAR-2001;

```
FEATURES             Location/Qualifiers
     source          1. .1026
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match           100.0%;  Score 1026;  DB 6;  Length 1026;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-179;
  Matches 1026;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTGAAGAACGAAAAAACCTTCTTTGAAGAGCTTTACGAGGCTTTAGAGGAAACCCACGAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTGAAGAACGAAAAAACCTTCTTTGAAGAGCTTTACGAGGCTTTAGAGGAAACCCACGAC 60

Qy         61 AACACCGATGCCACTAGGGGGTCAGATAGGGGGTCAGAGGACTTCTTCTTGGCCACCGAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACACCGATGCCACTAGGGGGTCAGATAGGGGGTCAGAGGACTTCTTCTTGGCCACCGAC 120

Qy        121 CCCCCTCCAGATGGAGGTGCCGAAAATCGCCTCGCGAAGGGCTTTACATACCAAAAAGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCCCTCCAGATGGAGGTGCCGAAAATCGCCTCGCGAAGGGCTTTACATACCAAAAAGAG 180

Qy        181 GCACTTAGGATTGCTTTACCCGAGAAAGACCATGAGGCTTTCCTTTCCTCTGTTGGGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GCACTTAGGATTGCTTTACCCGAGAAAGACCATGAGGCTTTCCTTTCCTCTGTTGGGGCC 240

Qy        241 CCCCCTATACCACCAGCTGAACCCCCCGTTGGGAATGTATGTCAAGCCGTCCAGGACGGG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCCCCTATACCACCAGCTGAACCCCCCGTTGGGAATGTATGTCAAGCCGTCCAGGACGGG 300

Qy        301 CCTCAGAAGCTTCTGGAACTCCTCCAGGAGATTGCCCGCTCCACCATCCCCTACGGCAAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CCTCAGAAGCTTCTGGAACTCCTCCAGGAGATTGCCCGCTCCACCATCCCCTACGGCAAC 360

Qy        361 CGGGAGCTCTGGAGGAAGGTGGGGACGGTCGTCTTCATGGTCCCCCTGGAGATGTTGGCC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CGGGAGCTCTGGAGGAAGGTGGGGACGGTCGTCTTCATGGTCCCCCTGGAGATGTTGGCC 420

Qy        421 CTCAACCTGGGGGTCACCCGGCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGAAAAAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTCAACCTGGGGGTCACCCGGCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGAAAAAG 480

Qy        481 GGCCTGGTGGCCACCGACGTCCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCATCGGC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GGCCTGGTGGCCACCGACGTCCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCATCGGC 540

Qy        541 ACCCTTTGGGCCGTCCGGCTGAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACTACATC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ACCCTTTGGGCCGTCCGGCTGAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACTACATC 600

Qy        601 TACCCCTGGAGGAACCTCGCCCTAGACATGGCCAACGGCGTGCTCTCCTTCAACTGGGTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TACCCCTGGAGGAACCTCGCCCTAGACATGGCCAACGGCGTGCTCTCCTTCAACTGGGTC 660

Qy        661 AAGGCCTACCAGGACCACGGAATCCGCCCCACCCTGGACGTGCTGGTCCTCTGGGCTCAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AAGGCCTACCAGGACCACGGAATCCGCCCCACCCTGGACGTGCTGGTCCTCTGGGCTCAG 720

Qy        721 GGGAAAAGGGTGATGCCCAACACCAAGACCGTGGCCGTTGACCTGGGCCTCATCCTGGTC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GGGAAAAGGGTGATGCCCAACACCAAGACCGTGGCCGTTGACCTGGGCCTCATCCTGGTC 780

Qy        781 CTCCCCGAGGTGGAGCGTTCCAAACTCCCGGCCCTTATCACCCTCATTGCTACGTACATT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CTCCCCGAGGTGGAGCGTTCCAAACTCCCGGCCCTTATCACCCTCATTGCTACGTACATT 840

Qy        841 GCCGATCTCCTAGATGACCGTCGTTCAAGACGTTTCTATGCAGGCTTGCTGTGGGCTGTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCCGATCTCCTAGATGACCGTCGTTCAAGACGTTTCTATGCAGGCTTGCTGTGGGCTGTG 900

Qy        901 GCCAGGGGTGAACTCCCCGCGCAATATCTATTTGCCGTCCTAATGCGGGTTATCCGAGAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GCCAGGGGTGAACTCCCCGCGCAATATCTATTTGCCGTCCTAATGCGGGTTATCCGAGAT 960

Qy        961 TACACGGATGGCCATCTGACACGACCGGGAGCGTACCTAGTGAAGACCCTCAAGGAGGCC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TACACGGATGGCCATCTGACACGACCGGGAGCGTACCTAGTGAAGACCCTCAAGGAGGCC 1020

Qy       1021 TCCTGA 1026
              ||||||
Db       1021 TCCTGA 1026


RESULT 2
U89376
LOCUS       U89376                  1600 bp    DNA     linear   BCT 06-OCT-1997
DEFINITION  Thermus sp. YS45 replication protein (repT) gene, complete cds.
ACCESSION   U89376
VERSION     U89376.1  GI:2465516
KEYWORDS    .
SOURCE      Thermus sp. YS45
  ORGANISM  Thermus sp. YS45
            Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;
            Thermus.
REFERENCE   1  (bases 1 to 1600)
  AUTHORS   Wayne,J. and Xu,S.Y.
  TITLE     Identification of a thermophilic plasmid origin and its cloning
            within a new Thermus-E. coli shuttle vector
  JOURNAL   Gene 195 (2), 321-328 (1997)
  MEDLINE   97449309
   PUBMED   9305778
REFERENCE   2  (bases 1 to 1600)
  AUTHORS   Wayne,J. and Xu,S.-Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-FEB-1997) Research, New England Biolabs, 32 Tozer
            Road, Beverly, MA 01915, USA
FEATURES             Location/Qualifiers
     source          1. .1600
                     /organism="Thermus sp. YS45"
                     /mol_type="genomic DNA"
                     /strain="YS45"
                     /db_xref="taxon:64227"
     -35_signal      220. .225
                     /note="putative"
     -10_signal      243. .248
                     /note="putative"
     -35_signal      278. .283
                     /note="putative"
     -10_signal      302. .307
                     /note="putative"
     gene            332. .1369
                     /gene="repT"
     RBS             332. .336
                     /gene="repT"
     CDS             344. .1369
                     /gene="repT"
                     /function="required for plasmid replication"
                     /note="RepT"
                     /codon_start=1
                     /transl_table=11
                     /product="replication protein"
                     /protein_id="AAB72144.1"
                     /db_xref="GI:2465517"
                     /translation="MKNEKTFFEELYEALEETHDNTDATRGSDRGSEDFFLATDPPPD
                     GGAENRLAKGFTYQKEALRIALPEKDHEAFLSSVGAPPIPPAEPPVGNVCQAVQDGPQ
                     KLLELLQEIARSTIPYGNRELWRKVGTVVFMVPLEMLALNLGVTRQTVHAWKKVLEKK
                     GLVATDVLHQTVNGERRAIGTLWAVRLRPGKARLTLDDYIYPWRNLALDMANGVLSFN
                     WVKAYQDHGIRPTLDVLVLWAQGKRVMPNTKTVAVDLGLILVLPEVERSKLPALITLI
                     ATYIADLLDDRRSRRFYAGLLWAVARGELPAQYLFAVLMRVIRDYTDGHLTRPGAYLV
                     KTLKEAS"
     misc_feature    1158. .1166
                     /gene="repT"
                     /note="putative; encodes DnaA box"
     misc_feature    1293. .1301
                     /gene="repT"
                     /note="putative; encodes DnaA box"
```

ORIGIN

Query Match 100.0%; Score 1026; DB 1; Length 1600;
Best Local Similarity 100.0%; Pred. No. 1.2e-179;
Matches 1026; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GTGAAGAACGAAAAAACCTTCTTTGAAGAGCTTTACGAGGCTTTAGAGGAAACCCACGAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        344 GTGAAGAACGAAAAAACCTTCTTTGAAGAGCTTTACGAGGCTTTAGAGGAAACCCACGAC 403

Qy         61 AACACCGATGCCACTAGGGGGTCAGATAGGGGGTCAGAGGACTTCTTCTTGGCCACCGAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        404 AACACCGATGCCACTAGGGGGTCAGATAGGGGGTCAGAGGACTTCTTCTTGGCCACCGAC 463

Qy        121 CCCCCTCCAGATGGAGGTGCCGAAAATCGCCTCGCGAAGGGCTTTACATACCAAAAAGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        464 CCCCCTCCAGATGGAGGTGCCGAAAATCGCCTCGCGAAGGGCTTTACATACCAAAAAGAG 523

Qy        181 GCACTTAGGATTGCTTTACCCGAGAAAGACCATGAGGCTTTCCTTTCCTCTGTTGGGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        524 GCACTTAGGATTGCTTTACCCGAGAAAGACCATGAGGCTTTCCTTTCCTCTGTTGGGGCC 583

Qy        241 CCCCCTATACCACCAGCTGAACCCCCCGTTGGGAATGTATGTCAAGCCGTCCAGGACGGG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        584 CCCCCTATACCACCAGCTGAACCCCCCGTTGGGAATGTATGTCAAGCCGTCCAGGACGGG 643

Qy        301 CCTCAGAAGCTTCTGGAACTCCTCCAGGAGATTGCCCGCTCCACCATCCCCTACGGCAAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        644 CCTCAGAAGCTTCTGGAACTCCTCCAGGAGATTGCCCGCTCCACCATCCCCTACGGCAAC 703

Qy        361 CGGGAGCTCTGGAGGAAGGTGGGGACGGTCGTCTTCATGGTCCCCCTGGAGATGTTGGCC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        704 CGGGAGCTCTGGAGGAAGGTGGGGACGGTCGTCTTCATGGTCCCCCTGGAGATGTTGGCC 763

Qy        421 CTCAACCTGGGGGTCACCCGGCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGAAAAAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        764 CTCAACCTGGGGGTCACCCGGCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGAAAAAG 823

Qy        481 GGCCTGGTGGCCACCGACGTCCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCATCGGC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        824 GGCCTGGTGGCCACCGACGTCCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCATCGGC 883

Qy        541 ACCCTTTGGGCCGTCCGGCTGAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACTACATC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        884 ACCCTTTGGGCCGTCCGGCTGAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACTACATC 943

Qy        601 TACCCCTGGAGGAACCTCGCCCTAGACATGGCCAACGGCGTGCTCTCCTTCAACTGGGTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        944 TACCCCTGGAGGAACCTCGCCCTAGACATGGCCAACGGCGTGCTCTCCTTCAACTGGGTC 1003

Qy        661 AAGGCCTACCAGGACCACGGAATCCGCCCCACCCTGGACGTGCTGGTCCTCTGGGCTCAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1004 AAGGCCTACCAGGACCACGGAATCCGCCCCACCCTGGACGTGCTGGTCCTCTGGGCTCAG 1063

Qy        721 GGGAAAAGGGTGATGCCCAACACCAAGACCGTGGCCGTTGACCTGGGCCTCATCCTGGTC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1064 GGGAAAAGGGTGATGCCCAACACCAAGACCGTGGCCGTTGACCTGGGCCTCATCCTGGTC 1123

Qy        781 CTCCCCGAGGTGGAGCGTTCCAAACTCCCGGCCCTTATCACCCTCATTGCTACGTACATT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1124 CTCCCCGAGGTGGAGCGTTCCAAACTCCCGGCCCTTATCACCCTCATTGCTACGTACATT 1183

Qy        841 GCCGATCTCCTAGATGACCGTCGTTCAAGACGTTTCTATGCAGGCTTGCTGTGGGCTGTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1184 GCCGATCTCCTAGATGACCGTCGTTCAAGACGTTTCTATGCAGGCTTGCTGTGGGCTGTG 1243

Qy        901 GCCAGGGGTGAACTCCCCGCGCAATATCTATTTGCCGTCCTAATGCGGGTTATCCGAGAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1244 GCCAGGGGTGAACTCCCCGCGCAATATCTATTTGCCGTCCTAATGCGGGTTATCCGAGAT 1303

Qy        961 TACACGGATGGCCATCTGACACGACCGGGAGCGTACCTAGTGAAGACCCTCAAGGAGGCC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1304 TACACGGATGGCCATCTGACACGACCGGGAGCGTACCTAGTGAAGACCCTCAAGGAGGCC 1363

Qy       1021 TCCTGA 1026
              ||||||
Db       1364 TCCTGA 1369
```

RESULT 3
AF013571/c
LOCUS AF013571 5849 bp DNA linear BCT 03-JAN-1998
DEFINITION Thermus sp. YS45 plasmid pTsp45s: Tsp45I restriction endonuclease (tsp45IR), Tsp45I methyltransferase (tsp45IM), and replication protein (repT) genes, complete cds.
ACCESSION AF013571
VERSION AF013571.1 GI:2735918
KEYWORDS .
SOURCE Thermus sp. YS45
ORGANISM Thermus sp. YS45
Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae; Thermus.
REFERENCE 1 (bases 1 to 5849)
AUTHORS Wayne,J. and Xu,S.Y.
TITLE Identification of a thermophilic plasmid origin and its cloning within a new Thermus-E. coli shuttle vector
JOURNAL Gene 195 (2), 321-328 (1997)
MEDLINE 97449309
PUBMED 9305778
REFERENCE 2 (bases 1 to 5849)
AUTHORS Wayne,J., Holden,M. and Xu,S.Y.
TITLE The Tsp45I restriction-modification system is plasmid-borne within its thermophilic host
JOURNAL Gene 202 (1-2), 83-88 (1997)
MEDLINE 98087421
PUBMED 9427549
REFERENCE 3 (bases 1 to 5849)
AUTHORS Wayne,J., Holden,M. and Xu,S.-Y.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-1997) Research, New England Biolabs, 32 Tozer Road, Beverly, MA 01915, USA
FEATURES Location/Qualifiers

```
     source          1. .5849
                     /organism="Thermus sp. YS45"
                     /mol_type="genomic DNA"
                     /strain="YS45"
                     /db_xref="taxon:64227"
                     /plasmid="pTsp45s"
     gene            <602. .>1612
                     /gene="tsp45IR"
     RBS             602. .606
                     /gene="tsp45IR"
     CDS             623. .1612
                     /gene="tsp45IR"
                     /note="R Tsp45I"
                     /codon_start=1
                     /transl_table=11
                     /product="Tsp45I restriction endonuclease"
                     /protein_id="AAC46043.1"
                     /db_xref="GI:2735919"
                     /translation="MAEWNVWTQRSVELLEKGYLDKLLQVYKGESGSSRSVPEEVEEK
                     LREAYKAYEGRQDSPEAETKLVEAVLNARKKVERSPFNHPYLPLVYYLVSEKAEKANK
                     ALEEALQEVASKHPETIRVLAKEAQRRGVEALIQRLKEPPEINRQIGPMFKRWYKEEL
                     KGKIEERLPGPTKPKIVVVSPEKSKPEQAPLIAEREAGIIIYTGSDEALKDAAKENLG
                     LGEEAELGTKGVDFYVVIRRSPEETWHLTGEVKFQSDFGGNQDNQKLVAKASIRLDLE
                     KRHIGIVVVDGMPVVSKFRGWAGLGKETIVTSVLLLPDLIAELYQKGEEALGL"
     gene            complement(<1609. .>2860)
                     /gene="tsp45IM"
     CDS             complement(1609. .2850)
                     /gene="tsp45IM"
                     /note="M Tsp45I"
                     /codon_start=1
                     /transl_table=11
                     /product="Tsp45I methyltransferase"
                     /protein_id="AAC46044.1"
                     /db_xref="GI:2735920"
```

```
                     /translation="MSRSYPGLTRKAPLKASKTSEPSPFRLVYPGKRDEKEILDQPTP
                     QLVLRKETLLFLGGNAPLFEIDPIGTYFLGENGQVLRWMLREPGGYAGKVQLVYIDPP
                     YGTGQQFLVGGDETDRVATVSQPKNGQLGYDDTLDGPQFVEFLRERLILLRELMADSG
                     LIFVHIDEKYGFEVKLILDEVFGRRNFVNHIARIASNPKNFSRKAFGSQKDMILVYSK
                     TRDYVWNESASPYSEEEIARLFPFVDENGERYTTNPLHAPGETKDGPTGRPWRGILPP
                     PGRHWRYPPEKLDELDAQGLIVWSKNGVPRKKVYARDRLKKGKKLQDVWQFKDPPYPR
                     YPTEKNLDMLKLIVQTGSNEGDLVLDPFAGSGTTLIASPLLKRRSIGIDASWEAVKAF
                     TRRVLEDFPRLQHKFEIVSAF"
     rep_origin      2164. .4500
                     /note="required for plasmid replication (putative)"
     RBS             complement(2856. .2860)
                     /gene="tsp45IM"
     gene            complement(<3132. .>4169)
                     /gene="repT"
     CDS             complement(3132. .4157)
                     /gene="repT"
                     /function="required for plasmid replication"
                     /note="RepT"
                     /codon_start=1
                     /transl_table=11
                     /product="replication protein"
                     /protein_id="AAC46045.1"
                     /db_xref="GI:2735921"
                     /translation="MKNEKTFFEELYEALEETHDNTDATRGSDRGSEDFFLATDPPPD
                     GGAENRLAKGFTYQKEALRIALPEKDHEAFLSSVGAPPIPPAEPPVGNVCQAVQDGPQ
                     KLLELLQEIARSTIPYGNRELWRKVGTVVFMVPLEMLALNLGVTRQTVHAWKKVLEKK
                     GLVATDVLHQTVNGERRAIGTLWAVRLRPGKARLTLDDYIYPWRNLALDMANGVLSFN
                     WVKAYQDHGIRPTLDVLVLWAQGKRVMPNTKTVAVDLGLILVLPEVERSKLPALITLI
                     ATYIADLLDDRRSRRFYAGLLWAVARGELPAQYLFAVLMRVIRDYTDGHLTRPGAYLV
                     KTLKEAS"
     misc_feature    complement(3200. .3208)
                     /gene="repT"
                     /note="putative DnaA box required for plasmid replication"
     misc_feature    complement(3335. .3343)
                     /gene="repT"
                     /note="putative DnaA box required for plasmid replication"
     RBS             complement(4165. .4169)
                     /gene="repT"
     -10_signal      complement(4194. .4199)
                     /note="putative"
     -35_signal      complement(4218. .4223)
                     /note="putative"
     -10_signal      complement(4253. .4258)
                     /note="putative"
     -35_signal      complement(4276. .4281)
                     /note="putative"
ORIGIN

  Query Match           100.0%;  Score 1026;  DB 1;  Length 5849;
  Best Local Similarity  100.0%;  Pred. No. 9.6e-180;
  Matches 1026;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTGAAGAACGAAAAAACCTTCTTTGAAGAGCTTTACGAGGCTTTAGAGGAAACCCACGAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4157 GTGAAGAACGAAAAAACCTTCTTTGAAGAGCTTTACGAGGCTTTAGAGGAAACCCACGAC 4098

Qy         61 AACACCGATGCCACTAGGGGGTCAGATAGGGGGTCAGAGGACTTCTTCTTGGCCACCGAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4097 AACACCGATGCCACTAGGGGGTCAGATAGGGGGTCAGAGGACTTCTTCTTGGCCACCGAC 4038

Qy        121 CCCCCTCCAGATGGAGGTGCCGAAAATCGCCTCGCGAAGGGCTTTACATACCAAAAAGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4037 CCCCCTCCAGATGGAGGTGCCGAAAATCGCCTCGCGAAGGGCTTTACATACCAAAAAGAG 3978

Qy        181 GCACTTAGGATTGCTTTACCCGAGAAAGACCATGAGGCTTTCCTTTCCTCTGTTGGGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3977 GCACTTAGGATTGCTTTACCCGAGAAAGACCATGAGGCTTTCCTTTCCTCTGTTGGGGCC 3918

Qy        241 CCCCCTATACCACCAGCTGAACCCCCCGTTGGGAATGTATGTCAAGCCGTCCAGGACGGG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3917 CCCCCTATACCACCAGCTGAACCCCCCGTTGGGAATGTATGTCAAGCCGTCCAGGACGGG 3858

Qy        301 CCTCAGAAGCTTCTGGAACTCCTCCAGGAGATTGCCCGCTCCACCATCCCCTACGGCAAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3857 CCTCAGAAGCTTCTGGAACTCCTCCAGGAGATTGCCCGCTCCACCATCCCCTACGGCAAC 3798

Qy        361 CGGGAGCTCTGGAGGAAGGTGGGGACGGTCGTCTTCATGGTCCCCCTGGAGATGTTGGCC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3797 CGGGAGCTCTGGAGGAAGGTGGGGACGGTCGTCTTCATGGTCCCCCTGGAGATGTTGGCC 3738

Qy        421 CTCAACCTGGGGGTCACCCGGCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGAAAAAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3737 CTCAACCTGGGGGTCACCCGGCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGAAAAAG 3678

Qy        481 GGCCTGGTGGCCACCGACGTCCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCATCGGC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3677 GGCCTGGTGGCCACCGACGTCCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCATCGGC 3618

Qy        541 ACCCTTTGGGCCGTCCGGCTGAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACTACATC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3617 ACCCTTTGGGCCGTCCGGCTGAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACTACATC 3558

Qy        601 TACCCCTGGAGGAACCTCGCCCTAGACATGGCCAACGGCGTGCTCTCCTTCAACTGGGTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3557 TACCCCTGGAGGAACCTCGCCCTAGACATGGCCAACGGCGTGCTCTCCTTCAACTGGGTC 3498

Qy        661 AAGGCCTACCAGGACCACGGAATCCGCCCCACCCTGGACGTGCTGGTCCTCTGGGCTCAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3497 AAGGCCTACCAGGACCACGGAATCCGCCCCACCCTGGACGTGCTGGTCCTCTGGGCTCAG 3438

Qy        721 GGGAAAAGGGTGATGCCCAACACCAAGACCGTGGCCGTTGACCTGGGCCTCATCCTGGTC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3437 GGGAAAAGGGTGATGCCCAACACCAAGACCGTGGCCGTTGACCTGGGCCTCATCCTGGTC 3378

Qy        781 CTCCCCGAGGTGGAGCGTTCCAAACTCCCGGCCCTTATCACCCTCATTGCTACGTACATT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3377 CTCCCCGAGGTGGAGCGTTCCAAACTCCCGGCCCTTATCACCCTCATTGCTACGTACATT 3318

Qy        841 GCCGATCTCCTAGATGACCGTCGTTCAAGACGTTTCTATGCAGGCTTGCTGTGGGCTGTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3317 GCCGATCTCCTAGATGACCGTCGTTCAAGACGTTTCTATGCAGGCTTGCTGTGGGCTGTG 3258

Qy        901 GCCAGGGGTGAACTCCCCGCGCAATATCTATTTGCCGTCCTAATGCGGGTTATCCGAGAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3257 GCCAGGGGTGAACTCCCCGCGCAATATCTATTTGCCGTCCTAATGCGGGTTATCCGAGAT 3198

Qy        961 TACACGGATGGCCATCTGACACGACCGGGAGCGTACCTAGTGAAGACCCTCAAGGAGGCC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3197 TACACGGATGGCCATCTGACACGACCGGGAGCGTACCTAGTGAAGACCCTCAAGGAGGCC 3138

Qy       1021 TCCTGA 1026
              ||||||
Db       3137 TCCTGA 3132
```

```
RESULT 4
AR139426/c
LOCUS       AR139426               5849 bp    DNA     linear   PAT 16-JUN-2001
DEFINITION  Sequence 6 from patent US 6207377.
ACCESSION   AR139426
VERSION     AR139426.1  GI:14481922
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 5849)
  AUTHORS   Wayne,J. and Xu,S.-y.
  TITLE     Method for construction of thermus-E. coli shuttle vectors and
            identification of two Thermus plasmid replication origins
  JOURNAL   Patent: US 6207377-A 6 27-MAR-2001;
FEATURES             Location/Qualifiers
     source          1. .5849
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN
```

```
  Query Match           100.0%;  Score 1026;  DB 6;  Length 5849;
  Best Local Similarity  100.0%;  Pred. No. 9.6e-180;
  Matches 1026;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTGAAGAACGAAAAAACCTTCTTTGAAGAGCTTTACGAGGCTTTAGAGGAAACCCACGAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4157 GTGAAGAACGAAAAAACCTTCTTTGAAGAGCTTTACGAGGCTTTAGAGGAAACCCACGAC 4098

Qy         61 AACACCGATGCCACTAGGGGGTCAGATAGGGGGTCAGAGGACTTCTTCTTGGCCACCGAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4097 AACACCGATGCCACTAGGGGGTCAGATAGGGGGTCAGAGGACTTCTTCTTGGCCACCGAC 4038

Qy        121 CCCCCTCCAGATGGAGGTGCCGAAAATCGCCTCGCGAAGGGCTTTACATACCAAAAAGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4037 CCCCCTCCAGATGGAGGTGCCGAAAATCGCCTCGCGAAGGGCTTTACATACCAAAAAGAG 3978

Qy        181 GCACTTAGGATTGCTTTACCCGAGAAAGACCATGAGGCTTTCCTTTCCTCTGTTGGGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3977 GCACTTAGGATTGCTTTACCCGAGAAAGACCATGAGGCTTTCCTTTCCTCTGTTGGGGCC 3918

Qy        241 CCCCCTATACCACCAGCTGAACCCCCCGTTGGGAATGTATGTCAAGCCGTCCAGGACGGG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3917 CCCCCTATACCACCAGCTGAACCCCCGTTGGGAATGTATGTCAAGCCGTCCAGGACGGG 3858

Qy        301 CCTCAGAAGCTTCTGGAACTCCTCCAGGAGATTGCCCGCTCCACCATCCCCTACGGCAAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3857 CCTCAGAAGCTTCTGGAACTCCTCCAGGAGATTGCCCGCTCCACCATCCCCTACGGCAAC 3798

Qy        361 CGGGAGCTCTGGAGGAAGGTGGGGACGGTCGTCTTCATGGTCCCCCTGGAGATGTTGGCC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3797 CGGGAGCTCTGGAGGAAGGTGGGGACGGTCGTCTTCATGGTCCCCCTGGAGATGTTGGCC 3738

Qy        421 CTCAACCTGGGGGTCACCCGGCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGAAAAAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3737 CTCAACCTGGGGGTCACCCGGCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGAAAAAG 3678

Qy        481 GGCCTGGTGGCCACCGACGTCCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCATCGGC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3677 GGCCTGGTGGCCACCGACGTCCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCATCGGC 3618

Qy        541 ACCCTTTGGGCCGTCCGGCTGAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACTACATC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3617 ACCCTTTGGGCCGTCCGGCTGAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACTACATC 3558

Qy        601 TACCCCTGGAGGAACCTCGCCCTAGACATGGCCAACGGCGTGCTCTCCTTCAACTGGGTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3557 TACCCCTGGAGGAACCTCGCCCTAGACATGGCCAACGGCGTGCTCTCCTTCAACTGGGTC 3498

Qy        661 AAGGCCTACCAGGACCACGGAATCCGCCCCACCCTGGACGTGCTGGTCCTCTGGGCTCAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3497 AAGGCCTACCAGGACCACGGAATCCGCCCCACCCTGGACGTGCTGGTCCTCTGGGCTCAG 3438

Qy        721 GGGAAAAGGGTGATGCCCAACACCAAGACCGTGGCCGTTGACCTGGGCCTCATCCTGGTC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3437 GGGAAAAGGGTGATGCCCAACACCAAGACCGTGGCCGTTGACCTGGGCCTCATCCTGGTC 3378

Qy        781 CTCCCCGAGGTGGAGCGTTCCAAACTCCCGGCCCTTATCACCCTCATTGCTACGTACATT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3377 CTCCCCGAGGTGGAGCGTTCCAAACTCCCGGCCCTTATCACCCTCATTGCTACGTACATT 3318

Qy        841 GCCGATCTCCTAGATGACCGTCGTTCAAGACGTTTCTATGCAGGCTTGCTGTGGGCTGTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3317 GCCGATCTCCTAGATGACCGTCGTTCAAGACGTTTCTATGCAGGCTTGCTGTGGGCTGTG 3258

Qy        901 GCCAGGGGTGAACTCCCCGCGCAATATCTATTTGCCGTCCTAATGCGGGTTATCCGAGAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3257 GCCAGGGGTGAACTCCCCGCGCAATATCTATTTGCCGTCCTAATGCGGGTTATCCGAGAT 3198

Qy        961 TACACGGATGGCCATCTGACACGACCGGGAGCGTACCTAGTGAAGACCCTCAAGGAGGCC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3197 TACACGGATGGCCATCTGACACGACCGGGAGCGTACCTAGTGAAGACCCTCAAGGAGGCC 3138

Qy       1021 TCCTGA 1026
              ||||||
Db       3137 TCCTGA 3132
```

```
RESULT 5
SCO939124
LOCUS       SCO939124              308050 bp    DNA     linear   BCT 11-FEB-2003
DEFINITION  Streptomyces coelicolor A3(2) complete genome; segment 21/29.
ACCESSION   AL939124 AL023797 AL031031 AL031035 AL031124 AL031225 AL031231
            AL031260 AL034447 AL035559 AL035569 AL355913 AL645882
VERSION     AL939124.1  GI:24413886
KEYWORDS    .
SOURCE      Streptomyces coelicolor A3(2)
  ORGANISM  Streptomyces coelicolor A3(2)
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1
  AUTHORS   Bentley,S.D., Chater,K.F., Cerdeno-Tarraga,A.M., Challis,G.L.,
            Thomson,N.R., James,K.D., Harris,D.E., Quail,M.A., Kieser,H.,
            Harper,D., Bateman,A., Brown,S., Chandra,G., Chen,C.W., Collins,M.,
            Cronin,A., Fraser,A., Goble,A., Hidalgo,J., Hornsby,T., Howarth,S.,
            Huang,C.H., Kieser,T., Larke,L., Murphy,L., Oliver,K., O'Neil,S.,
            Rabbinowitsch,E., Rajandream,M.A., Rutherford,K., Rutter,S.,
            Seeger,K., Saunders,D., Sharp,S., Squares,R., Squares,S.,
            Taylor,K., Warren,T., Wietzorrek,A., Woodward,J., Barrell,B.G.,
            Parkhill,J. and Hopwood,D.A.
  TITLE     Complete genome sequence of the model actinomycete Streptomyces
            coelicolor A3(2)
  JOURNAL   Nature 417 (6885), 141-147 (2002)
  MEDLINE   21996410
   PUBMED   12000953
REFERENCE   2  (bases 1 to 308050)
  AUTHORS   Bentley,S.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces
            sequencing team, Sanger Institute, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA E-mail: sdb@sanger.ac.uk
COMMENT     On or before Oct 26, 2002 this sequence version replaced
            gi:20520664, gi:20520751, gi:20520752, gi:20520665, gi:20520811,
            gi:20520756, gi:20520845, gi:20520760, gi:20520817, gi:20520818,
            gi:20520782.
FEATURES             Location/Qualifiers
     source          1. .308050
                     /organism="Streptomyces coelicolor A3(2)"
                     /mol_type="genomic DNA"
                     /strain="A3(2)"
                     /db_xref="taxon:100226"
     gene            250. .2490
                     /gene="SCO5495"
                     /note="synonym: SC8D9.07"
     CDS             250. .2490
                     /gene="SCO5495"
                     /note="SC8D9.07, possible membrane associated
                     phosphodiesterase, len: 746 aa; similar to many eg.
                     TR:O87376 (EMBL:AF052518) c-di-GMP phosphodiesterase A
                     from Acetobacter xylinus cdg2 operon (752 aa) fasta
                     scores; opt: 959, z-score: 1059.2, E(): 0, (31.7% identity
                     in 603 aa overlap). Appears to have two distinct regions
                     with the first 300aa containing several potential membrane
                     spanning hydrophobic regions and the remainder of the
                     protein showing similarity to the phosphodiesterase
                     mentioned above. Also similar to a nearby gene on the same
                     cosmid (SC8D9.22) fasta scores; opt: 736, z-score: 625.6,
                     E(): 5.3e-30, (32.6% identity in 729 aa overlap). Contains
                     TTA Leu codon, possible target for bldA regulation. Also
                     contains Pfam match to entry PF00990 DUF9, Domain of
                     unknown function, score 166.70, E-value 3.9e-46 and Pfam
                     match to entry PF00563 DUF2, Domain of unknown function 2,
                     score 291.30, E-value 1.2e-83. Contains possible
                     hydrophobic membrane spanning regions"
                     /codon_start=1
                     /transl_table=11
                     /product="putative phosphodiesterase"
```

/protein_id="CAB37571.1"
/db_xref="GI:4467246"
/db_xref="GOA:Q9Z584"
/db_xref="SPTREMBL:Q9Z584"
/translation="MEPTESAAAGSRLRLRRMADAWHVSRWLGQRDGGVPSSMTGTRV
PQAVGRPHGHPQGHPHGLVDPDGERHLSWPALPAAVVATAAFVLGAGFYRAFSDGHAL
FPAGTVGWSLAVLSGIVVGHLVMLGRSRWWGGTGSGAALTLAVLLLYGWIPAGMVSLT
VVVLVGIARRGRWRQGVLHGAVDILGIAAGAVVLGVCGSVPSVERPWDPETWGVYAAP
KVVLVAVAYLAVTRALLWYIQTPRPGLPTAARTALVRHGLVAVALLGIAPLICVVAVA
KPLLLPLFAIPLIALDSTLWIARARAEEQLRDPLTGLPNRQWLLERTWTALDDAERIG
ARAALMLIDLDRFRSVNDTLGHLAGDRLLLQTADRLRQALPRGAEAARLGGDEFAVLL
PVADSTTSATRIARSLVGALSSPMDLDGLTLVLEASAGVAVFPDHALDAEGLLRRADV
AMYQAKRDRTGVEVYESKRDSNTPDRLGLLGDLRRALDAHEVQLHYQPKVRFDGQVAG
LEALVRWVHPERGKVPPDEFIAIAESSGLMPHLTEYVLETALGQVARWRAQGLFVPVA
VNVSPRDVHTPGFAGSVAARLARHGVPAGALQLEITEHVLLEDPQRAADTLNALTGHG
VKMSLDDFGTGYSSLVHLRRLPVSELKIDRSFVARLAIDNEDAEIVRCTVDLAHSLGL
LVVAEGVEDDETWERLRDLGCDAVQGWLVAAAMPPEETTAWLLARGSRGWVRAAAALP
AAASDE"
misc_feature 1141. .1635
/gene="SCO5495"
/note="Pfam match to entry PF00990 DUF9, Domain of unknown function, score 166.70, E-value 3.9e-46"
misc_feature 1420. .1422
/gene="SCO5495"
/note="TTA Leu codon, possible target for bldA regulation"
misc_feature 1672. .2406
/gene="SCO5495"
/note="Pfam match to entry PF00563 DUF2, Domain of unknown function 2, score 291.30, E-value 1.2e-83"
gene 2774. .3145
/gene="SCO5496"
/note="synonym: SC8D9.08"
CDS 2774. .3145
/gene="SCO5496"
/note="SC8D9.08, small hydrophobic membrane protein, len: 123 aa; no significant database similarities. Contains possible hydrophobic membrane spanning regions"
/codon_start=1
/transl_table=11
/product="putative small membrane hydrophobic protein"
/protein_id="CAB37572.1"
/db_xref="GI:4467247"
/db_xref="SPTREMBL:Q9Z583"
/translation="MTVTAYTLAVVLNLFCLFLGYRFLFQPASAATGYGVPADAGGDA
AAYLTVKGVRDGTLGVVGLALLAFAGARPEAWFMLCVALVPLADTLIVLRHGGTKAVA
FGIHFATAIVVLISAGLLFAV"
gene 3177. .3671
/gene="SCO5497"
/note="synonym: SC8D9.09"
CDS 3177. .3671
/gene="SCO5497"
/note="SC8D9.09, unknown, len: 164aa; similar to many of undefined function eg. TR:O06336 (EMBL:Z95390) hypothetical protein from Mycobacterium tuberculosis (177 aa) fasta scores; opt: 171, z-score: 213.4, E(): 0.00014, (34.9% identity in 109 aa overlap)."
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAB37573.1"
/db_xref="GI:4467248"
/db_xref="SPTREMBL:Q9Z582"
/translation="MSLFVPKFDESVVVREAEAEVVGRAPTTVRLLADSSSTGGALST
QRVTLTAGADGAKPHWHDNSAEMFFLLDGAADVLSGDEVLTAGPGDLIVVPPGKPHAF
AAVPGADADLLIVLAPGVERFEYFRHLQRIALGEATPESLLEVQELYDNHFLSSAAWD
ARRG"
gene 3816. .4112
/gene="SCO5498"
/note="synonyms: gatC, SC8D9.10"
CDS 3816. .4112
/gene="SCO5498"
/note="SC8D9.10, probable Glu-tRNA-Gln amidotransferase subunit C, gatC, len: 98aa; similar to many eg. TR:O06492 (EMBL:AF008553) Glu-tRNAGln amidotransferase subunit C, gatC, from Bacillus subtilis (101 aa) fasta scores; opt: 232. z-score: 294.8, E(): 4.1e-09, (42.1% identity in 95 aa overlap)."
/codon_start=1
/transl_table=11
/product="probable Glu-tRNAGln amidotransferase subunit C"
/protein_id="CAB37574.1"
/db_xref="GI:4467249"
/translation="MPGITREEVAHLARLARLELKPEELEHFAGQLDDIIGAVARVSE
VADQDVPPTSHPLPLTNVMRPDEVRPSLTPEQALSGAPAQEQQRFKVPQILGEE"
gene 4116. .5609
/gene="SCO5499"
/note="synonyms: gatA, SC8D9.11"
CDS 4116. .5609
/gene="SCO5499"
/note="SC8D9.11, probable Glu-tRNA-Gln amidotransferase subunit A, gatA, len: 497aa; similar to many eg. TR:O06491 (EMBL:AF008553) Glu-tRNAGln amidotransferase subunit A, gatA, from Bacillus subtilis (486 aa) fasta scores; opt: 1443, z-score: 1512.5, E(): 0, (52.0% identity in 465 aa overlap)."
/codon_start=1
/transl_table=11
/product="probable Glu-tRNA Gln amidotransferase subunit"
/protein_id="CAB37575.1"
/db_xref="GI:4467250"
/translation="MSDIIKLTAAEIAAKIASGELTAVQVTEAHLARIEAVDEKVHAF
LHVDREGALAQARAVDEKRERGEKLGPLAGVPLALKDIFTTEGIPTTVGSKILEGWIP
PYDATVTKRLKAADVVILGKTNMDEFAMGSSTENSAYGPTGNPWDLTKIPGGSGGGSS
AALAAFQAPLAIGTDTGGSIRQPASVTGTVGVKPTYGGVSRYGMVAFSSSLDQGGPCA
RTVLDAALLHEVIAGHDPMDSTSIDAPVPAVVEAARNGSVDGMRVGVVKQFRGEGYQA
GVVQRFDESVELLKSLGAEIVELDCPSFDLALSAYYLIAPSECSSNLARFDGLRYGAR
VGDDGTHSAEEVTSLTREAGFGPEVKRRIMLGTYALSSGYYDAYYGSAQKVRTLIKQE
FERAFEQVDVIVSPTTPTTAFAIGERADDPMAMYLADLCTIPTNLAGNAAMSLPCGLA
PEDNLPVGLQIIAPAMKDDRLYKVGAAVEAAFVEKWGHPLIEEAPSL"
gene 5606. .5845
/gene="SCO5500"
/note="synonym: SC8D9.12"
CDS 5606. .5845
/gene="SCO5500"
/note="SC8D9.12, putative membrane protein, len: 79 aa. Contains possible hydrophobic membrane spanning regions"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="CAB37576.1"
/db_xref="GI:4467251"
/db_xref="SPTREMBL:Q9Z579"
/translation="MSALTKAKGFKKSKSGLYVSMATSAFGAVGVYKQIKKARGESDT
LRLLDAVVTGAAVVTNLAILYRELKRLGDDDVLLG"
gene 5861. .7375
/gene="SCO5501"
/note="synonyms: gatB, SC8D9.13"
CDS 5861. .7375
/gene="SCO5501"
/note="SC8D9.13, probable Glu-tRNA-Gln amidotransferase subunit B, gatB, len: 504aa; similar to many eg. TR:O30509 (EMBL:AF008553) Glu-tRNA-Gln amidotransferase subunit B, gatB, from Bacillus subtilis (476 aa) fasta scores; opt:

Query Match 5.8%; Score 59.8; DB 1; Length 308050;
Best Local Similarity 46.3%; Pred. No. 0.16;
Matches 196; Conservative 0; Mismatches 227; Indels 0; Gaps 0;

Qy 374 GGAAGGTGGGGACGGTCGTCTTCATGGTCCCCCTGGAGATGTTGGCCCTCAACCTGGGGG 433
Db 107969 GGGAGGCGCTGGAGGCCGAGGGCCTGGTCACCCGGCTGCTGCCCCGACTGGGAGCGGGTCC 108028

Qy 434 TCACCCGGCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGAAAAAGGGCCTGGTGGCCA 493
Db 108029 GCTGCCGCCCGCAGCGCAACGCCGTGCACCTGTGGACCGTGGACCGGCACCTGATCGAGA 108088

```
Qy        494 CCGACGTCCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCATCGGCACCCTTTGGGCCG 553
              ||| |||||    |       ||| | |   |   |   || |   ||   |  ||
Db     108089 CCGCCGTCCGCGCCGCCGGGTTCACCCGCCGGGTGCACCGCCCCGACCTGCTCCTGATCG 108148

Qy        554 TCCGGCTGAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACTACATCTACCCCTGGAGGA 613
              |  ||||  ||  |  |  | ||     |   |  |||| ||  |    || |   ||
Db     108149 CCGCGCTGCTGCACGACATCGGCAAAGGCTGGCCCGGCGACCACTCCGTGGCCGGCGAGA 108208

Qy        614 ACCTCGCCCTAGACATGGCCAACGGCGTGCTCTCCTTCAACTGGGTCAAGGCCTACCAGG 673
              | ||||||  ||| | |||  | || |   || |  |  |  || || |||| || ||
Db     108209 CCATCGCCCGGGACGTCGCCGCCCGCATCGGCTTCGACGGCGCGGACACGGCCGTCCTGG 108268

Qy        674 ACCACGGAATCCGCCCCACCCTGGACGTGCTGGTCCTCTGGGCTCAGGGGAAAAGGGTGA 733
              |  |    | |||| |  |||     |  |||     |   | | | |   ||
Db     108269 CCACCCTCGTACGCCACCACCTCCTGCTCGTGGAGACCGCCACCCGGCGCGACCTGGACG 108328

Qy        734 TGCCCAACACCAAGACCGTGGCCGTTGACCTGGGCCTCATCCTGGTCCTCCCCGAGGTGG 793
              ||   ||||  |  || || ||   |  || |  || | |  |  |   ||| ||
Db     108329 ATCCGGCCACCGTGCGCGCGGTCGCGCAGGCGGTCGGCACCGAGCACACGCTGGAGCTGC 108388

Qy        794 AGC 796
              ||
Db     108389 TGC 108391
```

RESULT 6
AX655393/c

```
LOCUS       AX655393               2000 bp    DNA     linear   PAT 22-MAR-2003
DEFINITION  Sequence 5263 from Patent WO03000898.
ACCESSION   AX655393
VERSION     AX655393.1  GI:29158207
KEYWORDS    .
SOURCE      Oryza sativa
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   Chang,H.S., Chen,W., Cooper,B., Glazebrook,J., Goff,S.A., Hou,Y.M.,
            Katagiri,F., Quan,S., Tao,Y., Whitham,S., Xie,Z., Zhu,T. and Zou,G.
  TITLE     Plant genes involved in defense against pathogens
  JOURNAL   Patent: WO 03000898-A 5263 03-JAN-2003;
            Syngenta Participations AG (CH)
FEATURES             Location/Qualifiers
     source          1. .2000
                     /organism="Oryza sativa"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:4530"
ORIGIN
```

Query Match 5.5%; Score 56.8; DB 6; Length 2000;
Best Local Similarity 9.1%; Pred. No. 1.4;
Matches 60; Conservative 306; Mismatches 293; Indels 1; Gaps 1;

```
Qy        170 ACCAAAAAGAGGCACTTAGGATTGCTTTACCCGAGAAAGACCATGAGGCTTTCCTTTCCT 229
              :: :|:: :::: |   ::: :      :::   :  :| | :: : :|  :: | ::
Db        679 RMTRARMSKRRKWAGASMKSCWMYWRGARSMWYSKYSCSAKCCKKTRYMTSSYMSTGMYG 620

Qy        230 CTGTTGGGGCCCCCCCTATACCACCAGCTGAACCCCCCGTTGGGAATGTATGTCA-AGCC 288
              ::: ::: :  : :::  :: | :: :::|    : : :: | ::::  :::::: ::::
Db        619 MYSSYKSMSWTSKMSYMGKMTCTMYTSMKGSTRRSKMGRWSGMSRMYMRWWKKMRKRKYM 560

Qy        289 GTCCAGGACGGGCCTCAGAAGCTTCTGGAACTCCTCCAGGAGATTGCCCGCTCCACCATC 348
              ::: ::   :: :|: :|   :||: :::::|  :  :   :: :  :   ::|::: :
Db        559 RYMKWKCTWRRCMCYRWGYTMYTTSRSRMMYTGRYKARYTSKRRYMWYKYRKYCWYYYYG 500

Qy        349 CCCTACGGCAACCGGGAGCTCTGGAGGAAGGTGGGGACGGTCGTCTTCATGGTCCCCCTG 408
              :::: :  ::  : :  :  |:    :|| :     :::::::::  : :  ::: : :
Db        499 MYMKCSYMMRYGYCKACKKCCYAMCWKAAYSGMMMYWYRKYSKWMRMSTKYMWSMWYKKC 440

Qy        409 GAGATGTTGGCCCTCAACCTGGGGGTCACCCGGCAGACCGTCCACGCCTGGAAGAAGGTC 468
              :  :| :| : |::: :|  |  ::: :: :   ::|::: :: :     :   : ::
Db        439 RSMKYGAKGCYGCKMWTYCSYGYMKWYTYMGSYKYSRCYKYMRMYMYKGWMYMMYYSAYS 380
```

```
Qy        469 CTTGAGAAAAAGGGCCTGGTGGCCACCGACGTCCTTCACCAAACCGTCAACGGGGAGCGC 528
              :  :    |   :: ::: | :::: : ::: : | :|::  :  : ::   ::    :
Db        379 SMMTWYYYYAKYWKYWYKRRGTMSWYGKSYKKKYCTWWCYMKCMRCYRWRKMMRKKTKYS 320

Qy        529 CGGGCCATCGGCACCCTTTGGGCCGTCCGGCTGAGGCCAGGGAAAGCCAGGCTCACCCTG 588
              :  | ::   ::|||: ::| ::: :: :  |  ::   ::: ::::: ::: :  ::
Db        319 KRCYCWRYATCYWCCCYRKRGWYSRRSMMRTAGKWKMRSWSRWCRSYSWYKMYKKMWKKS 260

Qy        589 GACGACTACATCTACCCCTGGAGGAACCTCGCCCTAGACATGGCCAACGGCGTGCTCTCC 648
              ::  :|    |:: :  ::  :  :   :  : : :   ::: | : :: : :   ::|
Db        259 YYMSYGWARSSGTWSRSAAKRTYKGYSTSRRAKMMRACRMYSACRRYSRTSYYCGCSYCG 200

Qy        649 TTCAACTGGGTCAAGGCCTACCAGGACCACGGAATCCGCCCCACCCTGGACGTGCTGGTC 708
              : : ::: :::  |: :|:    |:     : || ::::: :|     |::  ::::
Db        199 SSKWKYMSKSCSMRMTCSSWCSCCYTCYYGAMCWSCCMSMMYMGSCGCYTRGWKWRSKYS 140

Qy        709 CTCTGGGCTCAGGGGAAAAGGGTGATGCCCAACACCAAGACCGTGGCCGTTGACCTGGGC 768
              : |::  : | :  :     :  :::::::  | :|   ::::: ::: :   :  ::
Db        139 MCCKKYCSCCTKYCSYTGYYRYCKWYKYSYYKCYYCYCYWYMSYMRYMMKCMCSRSCSSW 80

Qy        769 CTCATCCTGGTCCTCCCCGAGGTGGAGCGTTCCAAACTCCCGGCCCTTATCACCCTCATT 828
              : ||:|:|::|: ::::: | : |  |::  : :  :  ::::::: : |    || :::
Db         79 MSCAYCSTSSTSRWMSMYYAAKMGMCGSSGMYRMSKSCKMYSKYSSCKYTGSKKCTKRKY 20
```

RESULT 7
SCO939112/c

```
LOCUS       SCO939112              300800 bp    DNA     linear   BCT 11-FEB-2003
DEFINITION  Streptomyces coelicolor A3(2) complete genome; segment 9/29.
ACCESSION   AL939112 AL133220 AL133423 AL133471 AL136150 AL137166 AL137187
            AL163003 AL352972 AL356892 AL356992 AL357591 AL445327 AL645882
VERSION     AL939112.1  GI:24413764
KEYWORDS    .
SOURCE      Streptomyces coelicolor A3(2)
  ORGANISM  Streptomyces coelicolor A3(2)
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1
  AUTHORS   Bentley,S.D., Chater,K.F., Cerdeno-Tarraga,A.M., Challis,G.L.,
            Thomson,N.R., James,K.D., Harris,D.E., Quail,M.A., Kieser,H.,
            Harper,D., Bateman,A., Brown,S., Chandra,G., Chen,C.W., Collins,M.,
            Cronin,A., Fraser,A., Goble,A., Hidalgo,J., Hornsby,T., Howarth,S.,
            Huang,C.H., Kieser,T., Larke,L., Murphy,L., Oliver,K., O'Neil,S.,
            Rabbinowitsch,E., Rajandream,M.A., Rutherford,K., Rutter,S.,
            Seeger,K., Saunders,D., Sharp,S., Squares,R., Squares,S.,
            Taylor,K., Warren,T., Wietzorrek,A., Woodward,J., Barrell,B.G.,
            Parkhill,J. and Hopwood,D.A.
  TITLE     Complete genome sequence of the model actinomycete Streptomyces
            coelicolor A3(2)
  JOURNAL   Nature 417 (6885), 141-147 (2002)
  MEDLINE   21996410
   PUBMED   12000953
REFERENCE   2  (bases 1 to 300800)
  AUTHORS   Bentley,S.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces
            sequencing team, Sanger Institute, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA E-mail: sdb@sanger.ac.uk
COMMENT     On or before Oct 26, 2002 this sequence version replaced
            gi:20520847, gi:20520774, gi:20520824, gi:20520691, gi:20520851,
            gi:20520775, gi:20520830, gi:20520831, gi:20520832, gi:20520833,
            gi:20520835, gi:20520841.
FEATURES             Location/Qualifiers
     source          1. .300800
                     /organism="Streptomyces coelicolor A3(2)"
                     /mol_type="genomic DNA"
                     /strain="A3(2)"
                     /db_xref="taxon:100226"
     RBS             84. .89
     gene            98. .973
                     /gene="SCO2256"
                     /note="synonym: panB"
```

```
CDS             98. .973
                /gene="SCO2256"
                /note="SCC75A.02, panB, 3-methyl-2-oxobutanoate
                hydroxymethyltransferase (EC 2.1.2.11), len: 291 aa.
                Highly similar to many including: Escherichia coli
                SW:PANB_ECOLI(EMBL:L17086) 3-methyl-2-oxobutanoate
                hydroxymethyltransferase (EC 2.1.2.11) (264 aa), fasta
                scores opt: 711 z-score: 839.4 E(): 0 44.3% identity in
                262 aa overlap and Mycobacterium tuberculosis
                SW:PANB_MYCTU(EMBL:Z70692) 3-methyl-2-oxobutanoate
                hydroxymethyltransferase (EC 2.1.2.11) (281 aa), fasta
                scores opt: 1019 z-score: 1199.2 E(): 0 59.1% identity in
                269 aa overlap."
                /codon_start=1
                /transl_table=11
                /product="3-methyl-2-oxobutanoate hydroxymethyltransferase
                (EC 2.1.2.11)."
                /protein_id="CAB61702.1"
                /db_xref="GI:6469241"
                /db_xref="GOA:Q9RKS2"
                /db_xref="SWISS-PROT:Q9RKS2"
                /translation="MTQLSAAQTPQPKPADGNRALYGGKINRRITVRDITAAKERGEK
                WPMLTAYDAMTASVFDESGIPVMLVGDSAGNCHLGYETTVPVTLDEMTMLSAAVVRGT
                SRALIVGDLPFGSYQEGPVQALRSATRLVKEAGVGAVKLEGGERSHRQIELLVESGIP
                VMAHIGLTPQSVNSMGYRVQGRGEEAAQQLLRDAKAVQDAGAFAVVLELVPAELAAEV
                TRTLHIPTVGIGAGPDTDAQVLVWTDMLGLTGGRMPKFVKQYADLRKVMGDAAKAYAE
                DVVGGTFPADEHSVH"
gene            1093. .2190
                /gene="SCO2257"
                /note="synonym: SCC75A.03"
CDS             1093. .2190
                /gene="SCO2257"
                /note="SCC75A.03, probable ABC transporter, ATP-binding
                component, len: 365 aa. Highly similar to many ABC
                transport proteins e.g. Streptomyces peucetius
                SW:DRRA_STRPE(EMBL:M73758) daunorubicin resistance
                ATP-binding protein (330 aa), fasta scores opt: 922
                z-score: 1009.0 E():0 48.9% identity in 329 aa overlap and
                Streptomyces antibioticus TR:Q53716(EMBL:L06249)
                ATP-binding protein OLEC-ORF4 involved in
                oleandomycin-resistance (325 aa), fasta scores opt: 1883
                z-score: 2052.8 E():0 89.7% identity in 329 aa overlap.
                Also similar to several Streptomyces coelicolor putative
                ABC transporter ATP-binding components e.g
                TR:CAB42771(EMBL:AL049841) SCE9.25C (315 aa), fasta scores
                opt: 1031 z-score: 1127.7 E():0 54.4% identity in 318 aa
                overlap. Contains Prosite hits to PS00017 ATP/GTP-binding
                site motif A (P-loop) and PS00211 ABC transporters family
                signature. Also contains a Pfam match to entry PF00005
                ABC_tran, ABC transporter."
                /codon_start=1
                /transl_table=11
                /product="probable ABC transporter, ATP-binding
                component."
                /protein_id="CAB61703.1"
                /db_xref="GI:6469242"
                /db_xref="GOA:Q9RKS1"
                /db_xref="SPTREMBL:Q9RKS1"
                /translation="MAPAGCRWDVGALSVPPAIVWGMKRIDDIPRAADSAVTVRGLVK
                HYGETKALDGVDLDVREGTVMGVLGPNGAGKTTLVRILSTLITPDSGQAHVAGYDVVR
                QPRQLRRVIGLTGQYASVDEKLPGWENLYMIGRLLDLSRRDARTRADELLERFSLTEA
                GKRPASTYSGGMRRRLDLAASMIGHPAVLFLDEPTTGLDPRTRNEVWDEVKRMVGDGV
                TVLLTTQYMEEAEQLASELTVVDRGKVIANGGIEELKAKVGGRTLRIRPADPLQLRPL
                AAYLDELGITGLASSTVDTERGTVLVPILSDEQLTAVVGAVTARGITVSAVTTELPSL
                DEVFLSLTGHRASAPQDPAPVTDDTREEVAV"
misc_feature    1276. .1821
                /gene="SCO2257"
                /note="Pfam match to entry PF00005 ABC_tran, ABC
                transporter, score 189.70, E-value 4.5e-53"
misc_feature    1297. .1320
                /gene="SCO2257"
                /note="PS00017 ATP/GTP-binding site motif A (P-loop)"
misc_feature    1594. .1638
                /gene="SCO2257"
                /note="PS00211 ABC transporters family signature"
RBS             2174. .2180
                /gene="SCO2257"
gene            2187. .3011
                /gene="SCO2258"
                /note="synonym: SCC75A.04"
CDS             2187. .3011
                /gene="SCO2258"
                /note="SCC75A.04, probable ABC-transporter, transmembrane
                component, len: 274 aa. Similar to many transport proteins
                including: Streptomyces peucetius
                SW:DRRB_STRPE(EMBL:M73758) daunorubicin resistance
                transmembrane protein (283 aa), fasta scores opt: 523
                z-score: 593.2 E(): 1.2e-25 35.1% identity in 245 aa
                overlap and Streptomyces antibioticus
                TR:Q53717(EMBL:L06249) membrane protein OLEC-ORF5 involved
                in oleandomycin-resistance (273 aa), fasta scores opt:
                1502 z-score: 1685.2 E():0 83.2% identity in 274 aa
                overlap. Contains multiple possible membrane spanning
                hydrophobic domains."
                /codon_start=1
                /transl_table=11
                /product="probable ABC-transporter, transmembrane
                component."
                /protein_id="CAB61704.1"
                /db_xref="GI:6469243"
                /db_xref="GOA:Q9RKS0"
                /db_xref="SPTREMBL:Q9RKS0"
                /translation="MSAATYAPDTLTTPDARITPRAHLRHTSALIRRNLLWIRQDPES
                MFDAILFPIVFTLLFVYVFGGSIGQSLGGGQQAYVQYVIPGLMAMMGMNMAQGVGTGF
                NQDFNSGVMDRFRSLPIGRGSVLFAKIAVELLRMLFATAVLMVVAVLVGFDITDWAGL
                LASVGLSAVFGSALMWVFLTLGVVMKNAQSVQAMGFLVLMPLQFGSSIFAPTDSMPGW
                LQSFTDYNPLSSLADTARGLMLGGPVAHDLWVTLAWSAALTAVMAPIAIHKFRTKS"
gene            complement(3033. .7037)
                /gene="SCO2259"
                /note="synonym: SCC75A.05c"
CDS             complement(3033. .7037)
                /gene="SCO2259"
                /note="SCC75A.05c, possible multi-domain regulatory
                protein, len: 1334 aa. Similar in parts to many
                regulators. Similar in the N-terminus to Streptomyces
                coelicolor SW:AFSR_STRCO (EMBL:D90155) regulatory protein
                AfsR (993 aa), fasta scores opt: 524 z-score: 356.0 E():
                1.9e-12 37.0% identity in 316 aa overlap. Similar in the
                C-terminus to many other proposed regulators e.g.
                Mycobacterium tuberculosis TR:O53720 (EMBL:AL021931)
                transcriptional regulatory protein (1085 aa), fasta scores
                opt: 809 z-score: 541.8 E(): 8.3e-23 34.1% identity in 883
                aa overlap. Contains a Prosite hit to PS00017
                ATP/GTP-binding site motif A (P-loop). Also contains a
                highly degenerate repeat region with 16x PA(P/S)G and 5x
                APG. This repeat region also has a notably higher GC
                content (83.4%). Contains probable coiled-coils between
                residues 867 to 904 (38 residues) and 1093 to 1121 (29
                residues)."
                /codon_start=1
                /transl_table=11
                /product="putative multi-domain regulatory protein."
                /protein_id="CAB61705.1"
                /db_xref="GI:6469244"
                /db_xref="GOA:Q9RKR9"
                /db_xref="SPTREMBL:Q9RKR9"
                /translation="MTGSDGSGDNGPVRYRILGTTQVLRPDGTAVPLGGARLRALLTV
                LALRAGRAVPAGLLVEEVWAGDPPADATGALQALVGRLRRALGADSVASADGGYRLAA
                APDDVDLHRFDRLAGEGTRALADGDPAKAAVVLDDALALWRGPVLADLPDRTAEAARW
                ETRHFEALRARHTAALDLGQAEHSLPELTALCDGHPLDEPLQALRLRALRDSGRTAEA
                LAAYEAVRRLLADRLGTDPGPELRTLHAELLSPSPTPTPGRSRTPGWTSGPGPASGAG
                AASGTDVASGAGAASGPDPASGPASGPAVAPGSGGGPAPGWWPAPGTAPGSSTAPPHD
                TASAADTAPAPGPTSAPGTAPAAGTAAPAPGTAGPAPGTSYAPGTAPVAGTTPAPGTA
                PAPGTAGPARDTSYAPGTAPVAGTTPAPGTAPAPGSTPAPGTVPAPGTAPAPGPQPAD
                GRRPVTGPASGTGPGAATPPEAAAAASAGSAPSPAPEGPRPRGNLRARLTSFVGRDGD
                VERLRADLAATRLVTLLGPGGAGKTRLSQEAAEGAGDAARDGVWLAELAPVDDPADVP
```

```
                     EAVLTAVGARETVLYRAGAEEMRAATASEGQDTAVERLVEHCGRRRMLIVLDNCEHVV
                     DAAARLAEELLARCPRLTVLATSREPLGVPGESLRPVEPLPEPAALRLLADRGAAARP
                     GFRVDADEETAAACAEICRRLDGLPLAIELAAARLRMLTPRQIADRLDDRFRLLTSGS
                     RTVLPRQQTLRAVVDWSWDLLDADEREVLGRLSVFAGGCDLAAAEAVCGPAALDSLGS
                     LVDKSLVVAAPVTDRLSGDGMRYRLLETVAEYAGERLDEAGGRAAAARAHLTYYRELA
                     RTTDPLLRGPRQLAAIERLEREYENLRTALRHAVAERDEQEALCLALSLVWYWQMRDL
                     RVEARNWFVEVMALGPDPFAEPGTPARPVWERCTSAPPPMTGEVLAEARRGVHLAHLA
                     CMDTELDAWQNPAAQRKLRVIADTYEPGMPQTCSSPGLLWFYSVMLTGDMDRLRRIMD
                     ATVRTCRETPGYEWELAGGLQMRANMLANRTDWAGDAVRDADESLEIYDRLGDAWGMA

  Query Match             5.5%;  Score 56.8;  DB 1;  Length 300800;
  Best Local Similarity   45.6%;  Pred. No. 0.58;
  Matches  199;  Conservative    0;  Mismatches  237;  Indels    0;  Gaps    0;

Qy        339 CTCCACCATCCCCTACGGCAACCGGGAGCTCTGGAGGAAGGTGGGGACGGTCGTCTTCAT 398
Db     171393 CTCCGCCCTCGCCCAGCTGCCCGACCTCTACGTCAGCGAACTCAAGGCGCTGGTCTTCGG 171334

Qy        399 GGTCCCCCTGGAGATGTTGGCCCTCAACCTGGGGGTCACCCGGCAGACCGTCCACGCCTG 458
Db     171333 GTTCATCGCGGGCATCGTCGCCGCCTACCGGGGACTCAACCCGCGCGGCGGCCCCAAGGG 171274

Qy        459 GAAGAAGGTCCTTGAGAAAAAGGGCCTGGTGGCCACCGACGTCCTTCACCAAACCGTCAA 518
Db     171273 CGTCGGCGACGCCGTCAACCAGTCCGTCGTCATCACCTTCCTCCTGCTCTTCTTCGTCAA 171214

Qy        519 CGGGGAGCGCCGGGCCATCGGCACCCTTTGGGCCGTCCGGCTGAGGCCAGGGAAAGCCAG 578
Db     171213 CATGGTGATGACGGCCGTCTACCTCCAGATCGTCCGCCGAAGGGAGGCTGAGGTCCGAT 171154

Qy        579 GCTCACCCTGGACGACTACATCTACCCCTGGAGGAACCTCGCCCTAGACATGGCCAACGG 638
Db     171153 GGCCTCCCCGCTCGTCTGGCTCGACCGCTCCGGCGACCAACTGCTCTTCTACGTCCGGGC 171094

Qy        639 CGTGCTCTCCTTCAACTGGGTCAAGGCCTACCAGGACCACGGAATCCGCCCCACCCTGGA 698
Db     171093 CCTGCTGTGGGTCCCGCGGACCCTGCGCCGCTACCTCAAGGAGGTGCAGCGCCTCCTCGC 171034

Qy        699 CGTGCTGGTCCTCTGGGCTCAGGGGAAAAGGGTGATGCCCAACACCAAGACCGTGGCCGT 758
Db     171033 CGAGGTGGCCTTCGGCTCCGGCGGCCTCGGCGTCATCGGCGGCACCATCGGCGTGATGAT 170974

Qy        759 TGACCTGGGCCTCATC 774
Db     170973 CGCGATGACGCTCTTC 170958


RESULT 8
AX653795
LOCUS       AX653795               2030 bp    DNA     linear   PAT 22-MAR-2003
DEFINITION  Sequence 3665 from Patent WO03000898.
ACCESSION   AX653795
VERSION     AX653795.1  GI:29156609
KEYWORDS    .
SOURCE      Oryza sativa
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   Chang,H.S., Chen,W., Cooper,B., Glazebrook,J., Goff,S.A., Hou,Y.M.,
            Katagiri,F., Quan,S., Tao,Y., Whitham,S., Xie,Z., Zhu,T. and Zou,G.
  TITLE     Plant genes involved in defense against pathogens
  JOURNAL   Patent: WO 03000898-A 3665 03-JAN-2003;
            Syngenta Participations AG (CH)
FEATURES             Location/Qualifiers
     source          1. .2030
                     /organism="Oryza sativa"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:4530"
ORIGIN

  Query Match             5.5%;  Score 56.4;  DB 6;  Length 2030;
  Best Local Similarity   46.7%;  Pred. No. 1.6;
  Matches  214;  Conservative    0;  Mismatches  241;  Indels    3;  Gaps    1;

Qy        356 GCAACCGGGAGCTCTGGAGGAAGGTGGGGACGGTCGTCTTCATGGTCCCCCTGGAGATGT 415
Db        485 GCGCCATGGTGCTCACGCGGATGAAGGAGACGGCCGAGGCCTACCTCGGCGAGAAGGTCA 544

Qy        416 TGGCCCTCAACCTGGGGGTCACCCGGCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGA 475
Db        545 CGCGCGCCGTCGTCACCGTCCCGGCCTACTTCAACGACGCGCAGCGGCAGGCCACCAAGG 604

Qy        476 AAAAGGGCCTGGTGGCCACCGACGTCCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCA 535
Db        605 ACGCCGGCGTCATCGCCGGGCTCACCGTCGACCGCATCATCAACGAGCCGACCGCCGCCG 664

Qy        536 TCGGCACCCTTTGGGCCGTCCGGCTGAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACT 595
Db        665 CCATCGCCTACGGCATCGACAAGAAGGGCGCCGAGAAGAACGTCCTCGTCTTCGACCTCG 724

Qy        596 ACATCTACCCCTGGAGGAACCTCGCCCTAGACATGGCCAACGGCGTGCTCTCCTTCAACT 655
Db        725 GCGGCGGCACGTTCGACGTCAGCATCCTCGCCATCGACAACGGCGTGTTCGAGGTCCTTG 784

Qy        656 GGGTCAAGGCCTACCAGGACCACGGAATC---CGCCCCACCCTGGACGTGCTGGTCCTCT 712
Db        785 CCACCAACGGCGACACCCACCTCGGCGGCGAGGACTTCGACCAACGCCTCATGGACCACT 844

Qy        713 GGGCTCAGGGGAAAAGGGTGATGCCCAACACCAAGACCGTGGCCGTTGACCTGGGCCTCA 772
Db        845 TCGTCAAGGTCATCCGCCGGAAGCACGGGCGCGACATCGCCGGCGACGCGCGCGCTGG 904

Qy        773 TCCTGGTCCTCCCCGAGGTGGAGCGTTCCAAACTCCCG 810
Db        905 GCAAGCTCCGCCGCGAGTGCGAGCGCGCCAAGCGCGCG 942


RESULT 9
AX755741
LOCUS       AX755741               2031 bp    DNA     linear   PAT 23-JUN-2003
DEFINITION  Sequence 443 from Patent WO03000905.
ACCESSION   AX755741
VERSION     AX755741.1  GI:32167954
KEYWORDS    .
SOURCE      Oryza sativa
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   Zhu,T., Cheng,W., Briggs,S., Cooper,B., Goff,S.A., Moughamer,T.,
            Glazebrook,J., Katagiri,F., Kreps,J., Provart,N. and Ricke,D.
  TITLE     Identification and characterization of plant genes
  JOURNAL   Patent: WO 03000905-A 443 03-JAN-2003;
            Syngenta Participations AG (CH)
FEATURES             Location/Qualifiers
     source          1. .2031
                     /organism="Oryza sativa"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:4530"
ORIGIN

  Query Match             5.5%;  Score 56.4;  DB 6;  Length 2031;
  Best Local Similarity   46.7%;  Pred. No. 1.6;
  Matches  214;  Conservative    0;  Mismatches  241;  Indels    3;  Gaps    1;

Qy        356 GCAACCGGGAGCTCTGGAGGAAGGTGGGGACGGTCGTCTTCATGGTCCCCCTGGAGATGT 415
Db        485 GCGCCATGGTGCTCACGCGGATGAAGGAGACGGCCGAGGCCTACCTCGGCGAGAAGGTCA 544

Qy        416 TGGCCCTCAACCTGGGGGTCACCCGGCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGA 475
Db        545 CGCGCGCCGTCGTCACCGTCCCGGCCTACTTCAACGACGCGCAGCGGCAGGCCACCAAGG 604
```

```
Qy        476 AAAAGGGCCTGGTGGCCACCGACGTCCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCA 535
Db        605 ACGCCGGCGTCATCGCCGGGCTCACCGTCGACCGCATCATCAACGAGCCGACCGCCGCCG 664

Qy        536 TCGGCACCCTTTGGGCCGTCCGGCTGAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACT 595
Db        665 CCATCGCCTACGGCATCGACAAGAAGGGCGCCGAGAAGAACGTCCTCGTCTTCGACCTCG 724

Qy        596 ACATCTACCCCTGGAGGAACCTCGCCCTAGACATGGCCAACGGCGTGCTCTCCTTCAACT 655
Db        725 GCGGCGGCACGTTCGACGTCAGCATCCTCGCCATCGACAACGGCGTGTTCGAGGTCCTTG 784

Qy        656 GGGTCAAGGCCTACCAGGACCACGGAATC---CGCCCCACCCTGGACGTGCTGGTCCTCT 712
Db        785 CCACCAACGGCGACACCCACCTCGGCGGCGAGGACTTCGACCAACGCCTCATGGACCACT 844

Qy        713 GGGCTCAGGGGAAAAGGGTGATGCCCAACACCAAGACCGTGGCCGTTGACCTGGGCCTCA 772
Db        845 TCGTCAAGGTCATCCGCCGGAAGCACGGGCGCGACATCGCCGGCGACGCGCGCGCGCTGG 904

Qy        773 TCCTGGTCCTCCCCGAGGTGGAGCGTTCCAAACTCCCG 810
Db        905 GCAAGCTCCGCCGCGAGTGCGAGCGCGCCAAGCGCGCG 942


RESULT 10
AX660328
LOCUS       AX660328               2028 bp    DNA     linear   PAT 22-MAR-2003
DEFINITION  Sequence 685 from Patent WO03000906.
ACCESSION   AX660328
VERSION     AX660328.1  GI:29162163
KEYWORDS    .
SOURCE      Oryza sativa
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   Glazebrook,J., Briggs,S., Cooper,B., Goff,S.A., Moughamer,T.,
            Katagiri,F., Kreps,J., Provart,N., Ricke,D. and Zhu,T.
  TITLE     Plant disease resistance genes
  JOURNAL   Patent: WO 03000906-A 685 03-JAN-2003;
            Syngenta Participations AG (CH)
FEATURES             Location/Qualifiers
     source          1. .2028
                     /organism="Oryza sativa"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:4530"
ORIGIN

  Query Match             5.4%;  Score 55.4;  DB 6;  Length 2028;
  Best Local Similarity   46.5%;  Pred. No. 2.5;
  Matches  213;  Conservative    0;  Mismatches  242;  Indels    3;  Gaps    1;

Qy        356 GCAACCGGGAGCTCTGGAGGAAGGTGGGGACGGTCGTCTTCATGGTCCCCCTGGAGATGT 415
Db        485 GCGCCATGGTGCTCACCCGGATGAAGGAGACGGCCGAGGCCTACCTCGGCGAGAAGGTCA 544

Qy        416 TGGCCCTCAACCTGGGGGTCACCCGGCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGA 475
Db        545 CGCGCGCCGTCGTCACCGTCCCGGCCTACTTCAACGACGCGCAGCGGCAGGCCACCAAGG 604

Qy        476 AAAAGGGCCTGGTGGCCACCGACGTCCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCA 535
Db        605 ACGCCGGCGTCATCGCCGGGCTCACCGTCGACCGCATCATCAACGAGCCGACCGCCGCCG 664

Qy        536 TCGGCACCCTTTGGGCCGTCCGGCTGAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACT 595
Db        665 CCATCGCCTACGGCATCGACAAGAAGGGCGCNGAGAAGAACGTCCTCGTCTTCGACCTCG 724

Qy        596 ACATCTACCCCTGGAGGAACCTCGCCCTAGACATGGCCAACGGCGTGCTCTCCTTCAACT 655
Db        725 GCGGCGGCACGTTCGACGTCAGCATCCTCGCCATCGACAACGGCGTGTTCGAGGTCCTTG 784

Qy        656 GGGTCAAGGCCTACCAGGACCACGGAATC---CGCCCCACCCTGGACGTGCTGGTCCTCT 712
Db        785 CCACCAACGGCGACACCCACCTCGGCGGCGAGGACTTCGACCAACGCCTCATGGACCACT 844

Qy        713 GGGCTCAGGGGAAAAGGGTGATGCCCAACACCAAGACCGTGGCCGTTGACCTGGGCCTCA 772
Db        845 TCGTCAAGGTCATCCGCCGGAAGCACGGGCGCGACATCGCCGGCGACGCGCGCGCGCTGG 904

Qy        773 TCCTGGTCCTCCCCGAGGTGGAGCGTTCCAAACTCCCG 810
Db        905 GCAAGCTCCGCCGCGAGTGCGAGCGCGCCAAGCGCGCG 942


RESULT 11
AX755777
LOCUS       AX755777               2028 bp    DNA     linear   PAT 23-JUN-2003
DEFINITION  Sequence 517 from Patent WO03000905.
ACCESSION   AX755777
VERSION     AX755777.1  GI:32167972
KEYWORDS    .
SOURCE      Oryza sativa
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   Zhu,T., Cheng,W., Briggs,S., Cooper,B., Goff,S.A., Moughamer,T.,
            Glazebrook,J., Katagiri,F., Kreps,J., Provart,N. and Ricke,D.
  TITLE     Identification and characterization of plant genes
  JOURNAL   Patent: WO 03000905-A 517 03-JAN-2003;
            Syngenta Participations AG (CH)
FEATURES             Location/Qualifiers
     source          1. .2028
                     /organism="Oryza sativa"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:4530"
ORIGIN

  Query Match             5.4%;  Score 55.4;  DB 6;  Length 2028;
  Best Local Similarity   46.5%;  Pred. No. 2.5;
  Matches  213;  Conservative    0;  Mismatches  242;  Indels    3;  Gaps    1;

Qy        356 GCAACCGGGAGCTCTGGAGGAAGGTGGGGACGGTCGTCTTCATGGTCCCCCTGGAGATGT 415
Db        485 GCGCCATGGTGCTCACCCGGATGAAGGAGACGGCCGAGGCCTACCTCGGCGAGAAGGTCA 544

Qy        416 TGGCCCTCAACCTGGGGGTCACCCGGCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGA 475
Db        545 CGCGCGCCGTCGTCACCGTCCCGGCCTACTTCAACGACGCGCAGCGGCAGGCCACCAAGG 604

Qy        476 AAAAGGGCCTGGTGGCCACCGACGTCCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCA 535
Db        605 ACGCCGGCGTCATCGCCGGGCTCACCGTCGACCGCATCATCAACGAGCCGACCGCCGCCG 664

Qy        536 TCGGCACCCTTTGGGCCGTCCGGCTGAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACT 595
Db        665 CCATCGCCTACGGCATCGACAAGAAGGGCGCNGAGAAGAACGTCCTCGTCTTCGACCTCG 724

Qy        596 ACATCTACCCCTGGAGGAACCTCGCCCTAGACATGGCCAACGGCGTGCTCTCCTTCAACT 655
Db        725 GCGGCGGCACGTTCGACGTCAGCATCCTCGCCATCGACAACGGCGTGTTCGAGGTCCTTG 784

Qy        656 GGGTCAAGGCCTACCAGGACCACGGAATC---CGCCCCACCCTGGACGTGCTGGTCCTCT 712
Db        785 CCACCAACGGCGACACCCACCTCGGCGGCGAGGACTTCGACCAACGCCTCATGGACCACT 844

Qy        713 GGGCTCAGGGGAAAAGGGTGATGCCCAACACCAAGACCGTGGCCGTTGACCTGGGCCTCA 772
Db        845 TCGTCAAGGTCATCCGCCGGAAGCACGGGCGCGACATCGCCGGCGACGCGCGCGCGCTGG 904

Qy        773 TCCTGGTCCTCCCCGAGGTGGAGCGTTCCAAACTCCCG 810
Db        905 GCAAGCTCCGCCGCGAGTGCGAGCGCGCCAAGCGCGCG 942
```

```
Db        905 GCAAGCTCCGCCGCGAGTGCGAGCGCGCCAAGCGCGCG 942

RESULT 12
AX653794
LOCUS       AX653794                2028 bp    DNA     linear   PAT 22-MAR-2003
DEFINITION  Sequence 3664 from Patent WO03000898.
ACCESSION   AX653794
VERSION     AX653794.1  GI:29156608
KEYWORDS    .
SOURCE      Oryza sativa
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   Chang,H.S., Chen,W., Cooper,B., Glazebrook,J., Goff,S.A., Hou,Y.M.,
            Katagiri,F., Quan,S., Tao,Y., Whitham,S., Xie,Z., Zhu,T. and Zou,G.
  TITLE     Plant genes involved in defense against pathogens
  JOURNAL   Patent: WO 03000898-A 3664 03-JAN-2003;
            Syngenta Participations AG (CH)
FEATURES             Location/Qualifiers
     source          1..2028
                     /organism="Oryza sativa"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:4530"
ORIGIN

  Query Match              5.3%;  Score 54.8;  DB 6;  Length 2028;
  Best Local Similarity   46.5%;  Pred. No. 3.2;
  Matches  213;  Conservative    0;  Mismatches  242;  Indels    3;  Gaps    1;

Qy        356 GCAACCGGGAGCTCTGGAGGAAGGTGGGGACGGTCGTCTTCATGGTCCCCCTGGAGATGT 415
Db        485 GCGCCATGGTGCTCACCCGGATGAAGGAGACGGCCGAGGCCTACCTCGGCGAGAAGGTCA 544

Qy        416 TGGCCCTCAACCTGGGGGTCACCCGGCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGA 475
Db        545 CGCGCGCCGTCGTCACCGTCCCGGCCTACTTCAACGACGCGCAGCGGCAGGCCACCAAGG 604

Qy        476 AAAAGGGCCTGGTGGCCACCGACGTCCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCA 535
Db        605 ACGCCGGCGTCATCGCCGGGCTCACCGTCGACCGCATCATCAACGAGCCGACCGCCGCCG 664

Qy        536 TCGGCACCCTTTGGGCCGTCCGGCTGAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACT 595
Db        665 CCATCGCCTACGGCATCGACAAGAAGGGCGCSGAGAAGAACGTCCTCGTCTTCGACCTCG 724

Qy        596 ACATCTACCCCTGGAGGAACCTCGCCCTAGACATGGCCAACGGCGTGCTCTCCTTCAACT 655
Db        725 GCGGCGGCACGTTCGACGTCAGCATCCTCGCCATCGACAACGGCGTGTTCGAGGTCCTTG 784

Qy        656 GGGTCAAGGCCTACCAGGACCACGGAATCCG---CCCCACCCTGGACGTGCTGGTCCTCT 712
Db        785 CCACCAACGGCGACACCCACCTCGGCGGCGAGGACTTCGACCAACGCCTCATGGACCACT 844

Qy        713 GGGCTCAGGGGAAAAGGGTGATGCCCAACACCAAGACCGTGGCCGTTGACCTGGGCCTCA 772
Db        845 TCGTCAAGGTCATCCGCCGGAAGCACGGGCGCGACATCGCCGGCGACGCGCGCGCGCTGG 904

Qy        773 TCCTGGTCCTCCCCGAGGTGGAGCGTTCCAAACTCCCG 810
Db        905 GCAAGCTCCGCCGCGAGTGCGAGCGCGCCAAGCGCGCG 942

RESULT 13
AX654685
LOCUS       AX654685                2060 bp    DNA     linear   PAT 22-MAR-2003
DEFINITION  Sequence 4555 from Patent WO03000898.
ACCESSION   AX654685
VERSION     AX654685.1  GI:29157499
KEYWORDS    .
SOURCE      Oryza sativa
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   Chang,H.S., Chen,W., Cooper,B., Glazebrook,J., Goff,S.A., Hou,Y.M.,
            Katagiri,F., Quan,S., Tao,Y., Whitham,S., Xie,Z., Zhu,T. and Zou,G.
  TITLE     Plant genes involved in defense against pathogens
  JOURNAL   Patent: WO 03000898-A 4555 03-JAN-2003;
            Syngenta Participations AG (CH)
FEATURES             Location/Qualifiers
     source          1..2060
                     /organism="Oryza sativa"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:4530"
ORIGIN

  Query Match              5.3%;  Score 54.8;  DB 6;  Length 2060;
  Best Local Similarity   46.5%;  Pred. No. 3.2;
  Matches  213;  Conservative    0;  Mismatches  242;  Indels    3;  Gaps    1;

Qy        356 GCAACCGGGAGCTCTGGAGGAAGGTGGGGACGGTCGTCTTCATGGTCCCCCTGGAGATGT 415
Db        485 GCGCCATGGTGCTCACCCGGATGAAGGAGACGGCCGAGGCCTACCTCGGCGAGAAGGTCA 544

Qy        416 TGGCCCTCAACCTGGGGGTCACCCGGCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGA 475
Db        545 CGCGCGCCGTCGTCACCGTCCCGGCCTACTTCAACGACGCGCAGCGGCAGGCCACCAAGG 604

Qy        476 AAAAGGGCCTGGTGGCCACCGACGTCCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCA 535
Db        605 ACGCCGGCGTCATCGCCGGGCTCACCGTCGACCGCATCATCAACGAGCCGACCGCCGCCG 664

Qy        536 TCGGCACCCTTTGGGCCGTCCGGCTGAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACT 595
Db        665 CCATCGCCTACGGCATCGACAAGAAGGGCGCGGAGAAGAACGTCCTCGTCTTCGACCTCG 724

Qy        596 ACATCTACCCCTGGAGGAACCTCGCCCTAGACATGGCCAACGGCGTGCTCTCCTTCAACT 655
Db        725 GCGGCGGCACGTTCGACGTCAGCATCCTCGCCATCGACAACGGCGTGTTCGAGGTCCTTG 784

Qy        656 GGGTCAAGGCCTACCAGGACCACGGAATCCG---CCCCACCCTGGACGTGCTGGTCCTCT 712
Db        785 CCACCAACGGCGACACCCACCTCGGCGGCGAGGACTTCGACCAACGCCTCATGGACCACT 844

Qy        713 GGGCTCAGGGGAAAAGGGTGATGCCCAACACCAAGACCGTGGCCGTTGACCTGGGCCTCA 772
Db        845 TCGTCAAGGTCATCCGCCGGAAGCACGGGCGCGACATCGCCGGCGACGCGCGCGCGCTGG 904

Qy        773 TCCTGGTCCTCCCCGAGGTGGAGCGTTCCAAACTCCCG 810
Db        905 GCAAGCTCCGCCGCGAGTGCGAGCGCGCCAAGCGCGCG 942

RESULT 14
AK106696
LOCUS       AK106696                2273 bp    mRNA    linear   PLN 24-JUL-2003
DEFINITION  Oryza sativa (japonica cultivar-group) cDNA clone:002-114-D06, full
            insert sequence.
ACCESSION   AK106696
VERSION     AK106696.1  GI:32991905
KEYWORDS    FLI_CDNA; oligo capping.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   The Rice Full-Length cDNA Consortium, National Institute of
            Agrobiological Sciences Rice Full-Length cDNA Project Team:,
            Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K.,
            Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I.,
            Kojima,K., Namiki,T., Ohneda,E., Yahagi,W., Suzuki,K., Li,C.,
```

Ohtsuki,K., Shishiki,T., Foundation of Advancement of International Science Genome Sequencing & Analysis Group:, Otomo,Y., Murakami,K., Iida,Y., Sugano,S., Fujimura,T., Suzuki,Y., Tsunoda,Y., Kurosaki,T., Kodama,T., Masuda,H., Kobayashi,M., Xie,Q., Lu,M., Narikawa,R., Sugiyama,A., Mizuno,K., Yokomizo,S., Niikura,J., Ikeda,R., Ishibiki,J., Kawamata,M., Yoshimura,A., Miura,J., Kusumegi,T., Oka,M., Ryu,R., Ueda,M., Matsubara,K., RIKEN:, Kawai,J., Carninci,P., Adachi,J., Aizawa,K., Arakawa,T., Fukuda,S., Hara,A., Hashidume,W., Hayatsu,N., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kondo,S., Konno,H., Miyazaki,A., Osato,N., Ota,Y., Saito,R., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Yoshino,M. and Hayashizaki,Y.

TITLE Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice

JOURNAL Science 301 (5631), 376-379 (2003)

MEDLINE 22752273

PUBMED 12869764

REFERENCE 2 (bases 1 to 2273)

AUTHORS Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Doi,K., Fujimura,T., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayashizaki,Y., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Hotta,I., Iida,J., Iida,Y., Ikeda,R., Imamura,K., Imotani,K., Ishibiki,J., Ishii,Y., Ishikawa,M., Itoh,M., Kagawa,I., Kanagawa,S., Katoh,H., Kawagashira,N., Kawai,J., Kawamata,M., Kikuchi,S., Kishikawa-Hirozane,T., Kishimoto,N., Kobayashi,M., Kodama,T., Kojima,K., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Kurosaki,T., Kusumegi,T., Li,C., Lu,M., Masuda,H., Matsubara,K., Matsuyama,T., Miura,J., Miyazaki,A., Mizuno,K., Murakami,K., Murata,M., Nagata,T., Nakamura,M., Namiki,T., Narikawa,R., Niikura,J., Nishi,K., Nomura,K., Numasaki,R., Ohneda,E., Ohno,M., Ohtsuki,K., Oka,M., Ooka,H., Osato,N., Ota,Y., Otomo,Y., Ryu,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Satoh,K., Shibata,K., Shinagawa,A., Shiraki,T., Shishiki,T., Sogabe,Y., Sugano,S., Sugiyama,A., Suzuki,K., Suzuki,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Tsunoda,Y., Ueda,M., Waki,K., Xie,Q., Yahagi,W., Yamada,H., Yamamoto,M., Yasunishi,A., Yazaki,J., Yokomizo,S. and Yoshimura,A.

TITLE Direct Submission

JOURNAL Submitted (27-AUG-2002) Shoshi Kikuchi, National Institute of Agrobiological Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp, Tel:81-29-838-7007; Fax:81-29-838-7007)

COMMENT This clone is one of the 28K full-length cDNA clones from japonica rice.
URL : http://cdna01.dna.affrc.go.jp/cDNA/
NIAS Rice Full-Length cDNA Project Team: Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I., Kojima,K., Namiki,T., Ohneda,E., Yahagi,W., Suzuki,K., Li,C., Ohtsuki,K., Shishiki,T. and Yamamoto,M.
FAIS Genome Sequencing & Analysis Group: Otomo,Y., Iida,Y., Fujimura,T., Ikeda,R., Ishibiki,J., Kawamata,M., Kobayashi,M., Kodama,T., Kurosaki,T., Kusumegi,T., Lu,M., Masuda,H., Miura,J., Mizuno,K., Narikawa,R., Niikura,J., Oka,M., Ryu,R., Sugano,S., Sugiyama,A., Suzuki,Y., Tsunoda,Y., Ueda,M., Xie,Q., Yokomizo,S., Yoshimura,A., Matsubara,K. and Murakami,K.
Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken: Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Iida,J., Imamura,K., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kanagawa,S., Katoh,H., Kawai,J., Kishikawa-Hirozane,T., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Osato,N., Ota,Y., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Waki,K., Yasunishi,A. and Hayashizaki,Y.

FEATURES Location/Qualifiers
source 1. .2273
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="002-114-D06"

ORIGIN

Query Match 5.3%; Score 54.8; DB 8; Length 2273;
Best Local Similarity 46.5%; Pred. No. 3.1;
Matches 213; Conservative 0; Mismatches 242; Indels 3; Gaps 1;

```
Qy      356 GCAACCGGGAGCTCTGGAGGAAGGTGGGGACGGTCGTCTTCATGGTCCCCCTGGAGATGT 415
Db      554 GCGCCATGGTGCTCACCCGGATGAAGGAGACGGCCGAGGCCTACCTCGGCGAGAAGGTCA 613

Qy      416 TGGCCCTCAACCTGGGGGTCACCCGGCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGA 475
Db      614 CGCGCGCCGTCGTCACCGTCCCGGCCTACTTCAACGACGCGCAGCGGCAGGCCACCAAGG 673

Qy      476 AAAAGGGCCTGGTGGCCACCGACGTCCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCA 535
Db      674 ACGCCGGCGTCATCGCCGGGCTCACCGTCGACCGCATCATCAACGAGCCGACCGCCGCCG 733

Qy      536 TCGGCACCCTTTGGGCCGTCCGGCTGAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACT 595
Db      734 CCATCGCCTACGGCATCGACAAGAAGGGCGCGGAGAAGAACGTCCTCGTCTTCGACCTCG 793

Qy      596 ACATCTACCCCTGGAGGAACCTCGCCCTAGACATGGCCAACGGCGTGCTCTCCTTCAACT 655
Db      794 GCGGCGGCACGTTCGACGTCAGCATCCTCGCCATCGACAACGGCGTGTTCGAGGTCCTTG 853

Qy      656 GGGTCAAGGCCTACCAGGACCACGGAATCCG---CCCCACCCTGGACGTGCTGGTCCTCT 712
Db      854 CCACCAACGGCGACACCCACCTCGGCGGCGAGGACTTCGACCAACGCCTCATGGACCACT 913

Qy      713 GGGCTCAGGGGAAAAGGGTGATGCCCAACACCAAGACCGTGGCCGTTGACCTGGGCCTCA 772
Db      914 TCGTCAAGGTCATCCGCCGGAAGCACGGGCGCGACATCGCCGGCGACGCGCGCGCGCTGG 973

Qy      773 TCCTGGTCCTCCCCGAGGTGGAGCGTTCCAAACTCCCG 810
Db      974 GCAAGCTCCGCCGCGAGTGCGAGCGCGCCAAGCGCGCG 1011
```

RESULT 15
AC130600/c
LOCUS AC130600 137651 bp DNA linear PLN 25-NOV-2003
DEFINITION Oryza sativa (japonica cultivar-group) chromosome 5 clone OSJNBb0048I21, complete sequence.
ACCESSION AC130600
VERSION AC130600.2 GI:38502326
KEYWORDS HTG.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.

REFERENCE 1 (bases 1 to 137651)

AUTHORS Chow,T.-Y., Hsing,Y.-I.C., Chen,C.-S., Chen,H.-H., Liu,S.-M., Chao,Y.-T., Chang,S.-J., Chen,H.-C., Chen,S.-K., Chen,T.-R., Chen,Y.-L., Cheng,C.-H., Chung,C.-I., Han,S.-Y., Hsiao,S.-H., Hsiung,J.-N., Hsu,C.-H., Huang,J.-J., Kau,P.-I., Lee,M.-C., Leu,H.-L., Li,Y.-F., Lin,S.-J., Lin,Y.-C., Wu,S.-W., Yu,C.-Y., Yu,S.-W., Wu,H.-P. and Shaw,J.-F.

TITLE Oryza sativa BAC OSJNBb0048I21 genomic sequence

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 137651)

AUTHORS Chow,T.-Y. and Hsing,Y.-I.C.

TITLE Direct Submission

JOURNAL Submitted (13-AUG-2002) Institute of Botany, Academia Sinica, 128, Section 2, Academia Road, Nankang, Taipei 11529, Taiwan

```
REFERENCE   3  (bases 1 to 137651)
  AUTHORS   Chow,T.-Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-NOV-2003) Institute of Botany, Academia Sinica, 128,
            Section 2, Academia Road, Nankang, Taipei 11529, Taiwan
COMMENT     On Nov 25, 2003 this sequence version replaced gi:22212951.
FEATURES             Location/Qualifiers
     source          1. .137651
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="genomic DNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /chromosome="5"
                     /clone="OSJNBb0048I21"
ORIGIN

  Query Match             5.3%;  Score 54.8;  DB 8;  Length 137651;
  Best Local Similarity   46.5%;  Pred. No. 1.6;
  Matches  213;  Conservative    0;  Mismatches  242;  Indels    3;  Gaps    1;

Qy        356 GCAACCGGGAGCTCTGGAGGAAGGTGGGGACGGTCGTCTTCATGGTCCCCCTGGAGATGT 415
              ||  |  || ||||    ||| |  || ||||| ||    |    ||  |  | || |
Db      18020 GCGCCATGGTGCTCACCCGGATGAAGGAGACGGCCGAGGCCTACCTCGGCGAGAAGGTCA 17961

Qy        416 TGGCCCTCAACCTGGGGGTCACCCGGCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGA 475
              |  |  |  | |    ||| |      |   | | ||||   |  | ||| |    ||
Db      17960 CGCGCGCCGTCGTCACCGTCCCGGCCTACTTCAACGACGCGCAGCGGCAGGCCACCAAGG 17901

Qy        476 AAAAGGGCCTGGTGGCCACCGACGTCCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCA 535
              |    ||| |  | |||    |  | |   ||  | | |||||  |  |  |   |||
Db      17900 ACGCCGGCGTCATCGCCGGGCTCACCGTCGACCGCATCATCAACGAGCCGACCGCCGCCG 17841

Qy        536 TCGGCACCCTTTGGGCCGTCCGGCTGAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACT 595
              |  | ||    |   || |  |  | |  | | |||   |   |||  | | |||  |
Db      17840 CCATCGCCTACGGCATCGACAAGAAGGGCGCGGAGAAGAACGTCCTCGTCTTCGACCTCG 17781

Qy        596 ACATCTACCCCTGGAGGAACCTCGCCCTAGACATGGCCAACGGCGTGCTCTCCTTCAACT 655
              |  |  | | |        |  |  ||| | ||| | ||||||||||  ||    ||
Db      17780 GCGGCGGCACGTTCGACGTCAGCATCCTCGCCATCGACAACGGCGTGTTCGAGGTCCTTG 17721

Qy        656 GGGTCAAGGCCTACCAGGACCACGGAATCCG---CCCCACCCTGGACGTGCTGGTCCTCT 712
              ||| | | ||    ||| |||   |      |  |  ||    | |  ||| || ||
Db      17720 CCACCAACGGCGACACCCACCTCGGCGGCGAGGACTTCGACCAACGCCTCATGGACCACT 17661

Qy        713 GGGCTCAGGGGAAAAGGGTGATGCCCAACACCAAGACCGTGGCCGTTGACCTGGGCCTCA 772
              |   |||  |   |  || || |     | | | ||  | ||  |  |  |  ||
Db      17660 TCGTCAAGGTCATCCGCCGGAAGCACGGGCGCGACATCGCCGGCGACGCGCGCGCGCTGG 17601

Qy        773 TCCTGGTCCTCCCCGAGGTGGAGCGTTCCAAACTCCCG 810
              |  | ||| || ||||   |||||  |||| | | ||
Db      17600 GCAAGCTCCGCCGCGAGTGCGAGCGCGCCAAGCGCGCG 17563


Search completed: March 18, 2004, 02:39:39
Job time : 2614.35 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: March 17, 2004, 20:20:31 ; Search time 250.717 Seconds
(without alignments)
17384.739 Million cell updates/sec

Title: US-09-664-186-4
Perfect score: 1026
Sequence: 1 gtgaagaacgaaaaaacctt.....ccctcaaggaggcctcctga 1026

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1026	100.0	1026	5	AAD04666	Aad04666 Thermus r
c	2	1026	100.0	5849	5	AAD04668	Aad04668 Thermus p
c	3	56.8	5.5	2000	7	ADA71938	Ada71938 Rice gene
	4	56.4	5.5	2030	7	ADA70342	Ada70342 Rice gene
	5	56.4	5.5	2031	9	ADC08176	Adc08176 Rice DNA
	6	55.4	5.4	2028	8	ADA48615	Ada48615 Rice gene
	7	55.4	5.4	2028	9	ADC08212	Adc08212 Rice DNA
	8	54.8	5.3	2028	7	ADA70341	Ada70341 Rice gene
	9	54.8	5.3	2060	7	ADA71232	Ada71232 Rice gene
	10	54.4	5.3	615	2	AAQ37036	Aaq37036 SOD-T gen
	11	53.2	5.2	1755	7	ADA70624	Ada70624 Rice gene
	12	53.2	5.2	1965	8	ADA48223	Ada48223 Rice gene
	13	53.2	5.2	1965	9	ADC08174	Adc08174 Rice DNA
	14	52.4	5.1	113193	7	AAD54645	Aad54645 Streptomy
	15	51.6	5.0	1815	5	AAD09400	Aad09400 Zea mays
c	16	50.2	4.9	509	8	ACL23248	Acl23248 DNA clone
	17	50	4.9	2000	7	ADA71938	Ada71938 Rice gene
	18	49.6	4.8	1674	7	ACA38419	Aca38419 Prokaryot
	19	49.6	4.8	1677	7	ACA40670	Aca40670 Prokaryot
c	20	49.6	4.8	110000	4	AAI99682_27	Continuation (28 o
c	21	49.6	4.8	110000	4	AAI99683_27	Continuation (28 o
c	22	49.4	4.8	349980	6	ABQ81846	Abq81846 Bifidobac
	23	48.8	4.8	1236	9	ADC36241	Adc36241 Weed cont
	24	48.8	4.8	1454	9	ADC36242	Adc36242 Weed cont
c	25	48.6	4.7	21034	2	AAV62154	Aav62154 HSV-2 str
c	26	48.6	4.7	117213	2	AAV62176	Aav62176 HSV-2 str
c	27	48.6	4.7	154746	6	AAD25519	Aad25519 Human her
	28	47.8	4.7	963	7	ACA24030	Aca24030 Prokaryot
	29	47.2	4.6	1194	7	ADA69660	Ada69660 Rice gene
	30	47	4.6	6094	3	AAA71428	Aaa71428 Expressio
	31	46.8	4.6	1605	7	ABT31955	Abt31955 High-affi
	32	46.6	4.5	7659	7	ACD13414	Acd13414 Human DNA
	33	46.6	4.5	10065	7	AAL54465	Aal54465 Human cad
	34	46.6	4.5	10065	8	ACA63037	Aca63037 Human Cad
	35	46.6	4.5	10455	7	AAL54468	Aal54468 Human cad
	36	46.6	4.5	10455	8	ACA63040	Aca63040 Human Cad
	37	46.6	4.5	11073	7	AAL54466	Aal54466 Human cad
	38	46.6	4.5	11073	8	ACA63038	Aca63038 Human Cad
c	39	46.2	4.5	1473	2	AAV34317	Aav34317 Human sec
c	40	46.2	4.5	1473	6	AAD44668	Aad44668 Human sec
c	41	46.2	4.5	1473	6	AAD44886	Aad44886 Human sec
c	42	46.2	4.5	1473	7	ABX96998	Abx96998 Human sec
	43	46.2	4.5	1911	6	ABS78669	Abs78669 S. ghanae
	44	46.2	4.5	6236	7	ACC71561	Acc71561 VRC6802(p
	45	46	4.5	6800	6	ABK91541	Abk91541 Modified

ALIGNMENTS

RESULT 1
AAD04666
ID AAD04666 standard; DNA; 1026 BP.
XX
AC AAD04666;
XX
DT 04-JUL-2001 (first entry)
XX
DE Thermus replication protein, RepT encoding DNA from pTsp45S plasmid.
XX
KW Replication protein; RepT; partition protein; ParA; pTsp45S plasmid;
KW kanamycin-resistance gene; thermophilic transformation; Ori;
KW replication origin; ds.
XX
OS Thermus sp.
XX
FH Key Location/Qualifiers
FT CDS 1. .1026
FT /*tag= a
FT /product= "Replication protein, RepT"
FT /codon= (Seq:GTG, aa:Met)
XX
PN US6207377-B1.
XX
PD 27-MAR-2001.
XX
PF 14-AUG-1998; 98US-00134246.
XX
PR 14-AUG-1998; 98US-00134246.
XX
PA (NEWE) NEW ENGLAND BIOLABS INC.
XX
PI Wayne J, Xu S;
XX
DR WPI; 2001-298939/31.
DR P-PSDB; AAE01000.
XX
PT Cloning Thermus species (Ts) plasmid genes comprises transforming
PT Escherichia coli with cloned recombinant plasmid containing Ts and E.coli
PT origins of replication, isolating cloned recombinant plasmid from E.coli
PT and transforming Ts cell.
XX
PS Example I; Fig 1; 32pp; English.
XX
CC The present DNA sequence encodes Thermus sp. replication protein, RepT
CC which is obtained from pTsp45S plasmid. The replication protein, RepT is

```
CC   needed for thermophilic plasmid replication. The invention relates to
CC   Thermus sp. replication protein RepT, partition protein ParA and their
CC   corresponding DNA molecules which relates to recombinant DNA molecules
CC   encoding plasmid DNA replication origins in Thermus, as well as to
CC   shuttle vectors which contain the same. The invention also relates to
CC   method useful for cloning Thermus sp. plasmid genes which comprises
CC   inserting plasmid DNA comprising a Thermus sp. origin of replication
CC   (Ori) into a recombinant plasmid comprising a thermostable kanamycin-
CC   resistance gene and an Escherichia coli Ori, to produce a cloned
CC   recombinant plasmid. This cloned recombinant plasmid is transformed with
CC   an E. coli. host cell, and E. coli. host cell cultured for the expression
CC   of cloned recombinant plasmid. The cloned recombinant plasmid isolated
CC   from E. coli host cell is then transformed with Thermus sp. host cell and
CC   Thermus sp. host cell is cultured. Thus Thermus sp. plasmid genes are
CC   cloned. These plasmid DNAs are used for thermophilic transformation
XX
SQ   Sequence 1026 BP; 221 A; 323 C; 285 G; 197 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 1026;  DB 5;  Length 1026;
  Best Local Similarity   100.0%;  Pred. No. 5e-231;
  Matches 1026;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTGAAGAACGAAAAAACCTTCTTTGAAGAGCTTTACGAGGCTTTAGAGGAAACCCACGAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTGAAGAACGAAAAAACCTTCTTTGAAGAGCTTTACGAGGCTTTAGAGGAAACCCACGAC 60

Qy         61 AACACCGATGCCACTAGGGGGTCAGATAGGGGGTCAGAGGACTTCTTCTTGGCCACCGAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACACCGATGCCACTAGGGGGTCAGATAGGGGGTCAGAGGACTTCTTCTTGGCCACCGAC 120

Qy        121 CCCCCTCCAGATGGAGGTGCCGAAAATCGCCTCGCGAAGGGCTTTACATACCAAAAAGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCCCTCCAGATGGAGGTGCCGAAAATCGCCTCGCGAAGGGCTTTACATACCAAAAAGAG 180

Qy        181 GCACTTAGGATTGCTTTACCCGAGAAAGACCATGAGGCTTTCCTTTCCTCTGTTGGGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GCACTTAGGATTGCTTTACCCGAGAAAGACCATGAGGCTTTCCTTTCCTCTGTTGGGGCC 240

Qy        241 CCCCCTATACCACCAGCTGAACCCCCCGTTGGGAATGTATGTCAAGCCGTCCAGGACGGG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCCCCTATACCACCAGCTGAACCCCCCGTTGGGAATGTATGTCAAGCCGTCCAGGACGGG 300

Qy        301 CCTCAGAAGCTTCTGGAACTCCTCCAGGAGATTGCCCGCTCCACCATCCCCTACGGCAAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CCTCAGAAGCTTCTGGAACTCCTCCAGGAGATTGCCCGCTCCACCATCCCCTACGGCAAC 360

Qy        361 CGGGAGCTCTGGAGGAAGGTGGGGACGGTCGTCTTCATGGTCCCCCTGGAGATGTTGGCC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CGGGAGCTCTGGAGGAAGGTGGGGACGGTCGTCTTCATGGTCCCCCTGGAGATGTTGGCC 420

Qy        421 CTCAACCTGGGGGTCACCCGGCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGAAAAAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTCAACCTGGGGGTCACCCGGCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGAAAAAG 480

Qy        481 GGCCTGGTGGCCACCGACGTCCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCATCGGC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GGCCTGGTGGCCACCGACGTCCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCATCGGC 540

Qy        541 ACCCTTTGGGCCGTCCGGCTGAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACTACATC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ACCCTTTGGGCCGTCCGGCTGAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACTACATC 600

Qy        601 TACCCCTGGAGGAACCTCGCCCTAGACATGGCCAACGGCGTGCTCTCCTTCAACTGGGTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TACCCCTGGAGGAACCTCGCCCTAGACATGGCCAACGGCGTGCTCTCCTTCAACTGGGTC 660

Qy        661 AAGGCCTACCAGGACCACGGAATCCGCCCCACCCTGGACGTGCTGGTCCTCTGGGCTCAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AAGGCCTACCAGGACCACGGAATCCGCCCCACCCTGGACGTGCTGGTCCTCTGGGCTCAG 720

Qy        721 GGGAAAAGGGTGATGCCCAACACCAAGACCGTGGCCGTTGACCTGGGCCTCATCCTGGTC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GGGAAAAGGGTGATGCCCAACACCAAGACCGTGGCCGTTGACCTGGGCCTCATCCTGGTC 780

Qy        781 CTCCCCGAGGTGGAGCGTTCCAAACTCCCGGCCCTTATCACCCTCATTGCTACGTACATT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CTCCCCGAGGTGGAGCGTTCCAAACTCCCGGCCCTTATCACCCTCATTGCTACGTACATT 840

Qy        841 GCCGATCTCCTAGATGACCGTCGTTCAAGACGTTTCTATGCAGGCTTGCTGTGGGCTGTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCCGATCTCCTAGATGACCGTCGTTCAAGACGTTTCTATGCAGGCTTGCTGTGGGCTGTG 900

Qy        901 GCCAGGGGTGAACTCCCCGCGCAATATCTATTTGCCGTCCTAATGCGGGTTATCCGAGAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GCCAGGGGTGAACTCCCCGCGCAATATCTATTTGCCGTCCTAATGCGGGTTATCCGAGAT 960

Qy        961 TACACGGATGGCCATCTGACACGACCGGGAGCGTACCTAGTGAAGACCCTCAAGGAGGCC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TACACGGATGGCCATCTGACACGACCGGGAGCGTACCTAGTGAAGACCCTCAAGGAGGCC 1020

Qy       1021 TCCTGA 1026
              ||||||
Db       1021 TCCTGA 1026
```

```
RESULT 2
AAD04668/c
ID   AAD04668 standard; DNA; 5849 BP.
XX
AC   AAD04668;
XX
DT   04-JUL-2001  (first entry)
XX
DE   Thermus plasmid pTsp45S DNA sequence.
XX
KW   Replication protein; RepT; partition protein; ParA; pTsp45S plasmid;
KW   kanamycin-resistance gene; thermophilic transformation; Ori;
KW   replication origin; ds.
XX
OS   Thermus sp.
XX
PN   US6207377-B1.
XX
PD   27-MAR-2001.
XX
PF   14-AUG-1998;   98US-00134246.
XX
PR   14-AUG-1998;   98US-00134246.
XX
PA   (NEWE ) NEW ENGLAND BIOLABS INC.
XX
PI   Wayne J,  Xu S;
XX
DR   WPI; 2001-298939/31.
XX
PT   Cloning Thermus species (Ts) plasmid genes comprises transforming
PT   Escherichia coli with cloned recombinant plasmid containing Ts and E.coli
PT   origins of replication, isolating cloned recombinant plasmid from E.coli
PT   and transforming Ts cell.
XX
PS   Example I; Fig 3; 32pp; English.
XX
CC   The present sequence is Thermus plasmid pTsp45S DNA. The open reading
CC   frame of pTsp45S plasmid is the replication protein, RepT which is needed
CC   for thermophilic plasmid replication. The invention relates to Thermus
CC   sp. replication protein RepT, partition protein ParA and their
CC   corresponding DNA molecules which relates to recombinant DNA molecules
CC   encoding plasmid DNA replication origins in Thermus, as well as to
CC   shuttle vectors which contain the same. The invention also relates to
CC   method useful for cloning Thermus sp. plasmid genes which comprises
CC   inserting plasmid DNA comprising a Thermus sp. origin of replication
CC   (Ori) into a recombinant plasmid comprising a thermostable kanamycin-
CC   resistance gene and an Escherichia coli Ori, to produce a cloned
CC   recombinant plasmid. This cloned recombinant plasmid is transformed with
CC   an E. coli. host cell, and E. coli. host cell cultured for the expression
CC   of cloned recombinant plasmid. The cloned recombinant plasmid isolated
```

CC from E. coli host cell is then transformed with Thermus sp. host cell and
CC Thermus sp. host cell is cultured. Thus Thermus sp. plasmid genes are
CC cloned. These plasmid DNAs are used for thermophilic transformation
XX
SQ Sequence 5849 BP; 1377 A; 1501 C; 1670 G; 1301 T; 0 U; 0 Other;

Query Match 100.0%; Score 1026; DB 5; Length 5849;
Best Local Similarity 100.0%; Pred. No. 7.7e-231;
Matches 1026; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GTGAAGAACGAAAAAACCTTCTTTGAAGAGCTTTACGAGGCTTTAGAGGAAACCCACGAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4157 GTGAAGAACGAAAAAACCTTCTTTGAAGAGCTTTACGAGGCTTTAGAGGAAACCCACGAC 4098

Qy         61 AACACCGATGCCACTAGGGGGTCAGATAGGGGGTCAGAGGACTTCTTCTTGGCCACCGAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4097 AACACCGATGCCACTAGGGGGTCAGATAGGGGGTCAGAGGACTTCTTCTTGGCCACCGAC 4038

Qy        121 CCCCCTCCAGATGGAGGTGCCGAAAATCGCCTCGCGAAGGGCTTTACATACCAAAAAGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4037 CCCCCTCCAGATGGAGGTGCCGAAAATCGCCTCGCGAAGGGCTTTACATACCAAAAAGAG 3978

Qy        181 GCACTTAGGATTGCTTTACCCGAGAAAGACCATGAGGCTTTCCTTTCCTCTGTTGGGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3977 GCACTTAGGATTGCTTTACCCGAGAAAGACCATGAGGCTTTCCTTTCCTCTGTTGGGGCC 3918

Qy        241 CCCCCTATACCACCAGCTGAACCCCCCGTTGGGAATGTATGTCAAGCCGTCCAGGACGGG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3917 CCCCCTATACCACCAGCTGAACCCCCCGTTGGGAATGTATGTCAAGCCGTCCAGGACGGG 3858

Qy        301 CCTCAGAAGCTTCTGGAACTCCTCCAGGAGATTGCCCGCTCCACCATCCCCTACGGCAAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3857 CCTCAGAAGCTTCTGGAACTCCTCCAGGAGATTGCCCGCTCCACCATCCCCTACGGCAAC 3798

Qy        361 CGGGAGCTCTGGAGGAAGGTGGGGACGGTCGTCTTCATGGTCCCCCTGGAGATGTTGGCC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3797 CGGGAGCTCTGGAGGAAGGTGGGGACGGTCGTCTTCATGGTCCCCCTGGAGATGTTGGCC 3738

Qy        421 CTCAACCTGGGGGTCACCCGGCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGAAAAAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3737 CTCAACCTGGGGGTCACCCGGCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGAAAAAG 3678

Qy        481 GGCCTGGTGGCCACCGACGTCCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCATCGGC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3677 GGCCTGGTGGCCACCGACGTCCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCATCGGC 3618

Qy        541 ACCCTTTGGGCCGTCCGGCTGAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACTACATC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3617 ACCCTTTGGGCCGTCCGGCTGAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACTACATC 3558

Qy        601 TACCCCTGGAGGAACCTCGCCCTAGACATGGCCAACGGCGTGCTCTCCTTCAACTGGGTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3557 TACCCCTGGAGGAACCTCGCCCTAGACATGGCCAACGGCGTGCTCTCCTTCAACTGGGTC 3498

Qy        661 AAGGCCTACCAGGACCACGGAATCCGCCCCACCCTGGACGTGCTGGTCCTCTGGGCTCAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3497 AAGGCCTACCAGGACCACGGAATCCGCCCCACCCTGGACGTGCTGGTCCTCTGGGCTCAG 3438

Qy        721 GGGAAAAGGGTGATGCCCAACACCAAGACCGTGGCCGTTGACCTGGGCCTCATCCTGGTC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3437 GGGAAAAGGGTGATGCCCAACACCAAGACCGTGGCCGTTGACCTGGGCCTCATCCTGGTC 3378

Qy        781 CTCCCCGAGGTGGAGCGTTCCAAACTCCCGGCCCTTATCACCCTCATTGCTACGTACATT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3377 CTCCCCGAGGTGGAGCGTTCCAAACTCCCGGCCCTTATCACCCTCATTGCTACGTACATT 3318

Qy        841 GCCGATCTCCTAGATGACCGTCGTTCAAGACGTTTCTATGCAGGCTTGCTGTGGGCTGTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3317 GCCGATCTCCTAGATGACCGTCGTTCAAGACGTTTCTATGCAGGCTTGCTGTGGGCTGTG 3258

Qy        901 GCCAGGGGTGAACTCCCCGCGCAATATCTATTTGCCGTCCTAATGCGGGTTATCCGAGAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3257 GCCAGGGGTGAACTCCCCGCGCAATATCTATTTGCCGTCCTAATGCGGGTTATCCGAGAT 3198
```

```
Qy        961 TACACGGATGGCCATCTGACACGACCGGGAGCGTACCTAGTGAAGACCCTCAAGGAGGCC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3197 TACACGGATGGCCATCTGACACGACCGGGAGCGTACCTAGTGAAGACCCTCAAGGAGGCC 3138

Qy       1021 TCCTGA 1026
              ||||||
Db       3137 TCCTGA 3132
```

RESULT 3
ADA71938/c
ID ADA71938 standard; DNA; 2000 BP.
XX
AC ADA71938;
XX
DT 20-NOV-2003 (first entry)
XX
DE Rice gene, SEQ ID 5263.
XX
KW Plant; bacterial infection; fungal infection; viral infection; rice;
KW gene; ds.
XX
OS Oryza sativa.
XX
PN WO2003000898-A1.
XX
PD 03-JAN-2003.
XX
PF 22-JUN-2001; 2001WO-IB001105.
XX
PR 22-JUN-2001; 2001WO-IB001105.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX
DR WPI; 2003-175290/17.
XX
PT Identifying at least one gene involved in plant resistance or response to
PT pathogenic infection for conferring resistance or tolerance to a plant to
PT bacterial, fungal or viral infection by determining or detecting plant
PT gene expression.
XX
PS Claim 27; SEQ ID NO 5263; 899pp; English.
XX
CC The present invention relates to a method (M1) for identifying genes
CC involved in plant resistance or response to pathogenic infection. M1
CC comprises identifying a gene whose expression is significantly altered in
CC the incompatible interaction of plant gene expression relative to
CC expression of the gene in an uninfected plant, in a mutant plant that
CC does not express a gene associated with response to pathogenic infection,
CC or in a corresponding incompatible or compatible interaction. (M1) is
CC useful for conferring resistance to resistance or tolerance to a plant to
CC bacterial, fungal or viral infection. The present sequence was used to
CC illustrate the invention.
XX
SQ Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;

Query Match 5.5%; Score 56.8; DB 7; Length 2000;
Best Local Similarity 9.1%; Pred. No. 0.0019;
Matches 60; Conservative 306; Mismatches 293; Indels 1; Gaps 1;

```
Qy        170 ACCAAAAAGAGGCACTTAGGATTGCTTTACCCGAGAAAGACCATGAGGCTTTCCTTTCCT 229
              :: :|:: :::: |   ::: :      :::   :   :| | :: : :|  :: | ::
Db        679 RMTRARMSKRRKWAGASMKSCWMYWRGARSMWYSKYSCSAKCCKKTRYMTSSYMSTGMYG 620

Qy        230 CTGTTGGGGCCCCCCCCTATACCACCAGCTGAACCCCCCGTTGGGAATGTATGTCA-AGCC 288
              ::: ::: :  : :::  :: | :: :::|    : : :: | :::: :::::: ::::
Db        619 MYSSYKSMSWTSKMSYMGKMTCTMYTSMKGSTRRSKMGRWSGMSRMYMRWWKKMRKRKYM 560

Qy        289 GTCCAGGACGGGCCTCAGAAGCTTCTGGAACTCCTCCAGGAGATTGCCCGCTCCACCATC 348
```

```
                  ::: ::   :: :|: :|   :||: :::::|  :  :   :: :  :   ::|::: :
Db        559 RYMKWKCTWRRCMCYRWGYTMYTTSRSRMMYTGRYKARYTSKRRYMWYKYRKYCWYYYYG 500

Qy        349 CCCTACGGCAACCGGGAGCTCTGGAGGAAGGTGGGGACGGTCGTCTTCATGGTCCCCCTG 408
              :::: :  ::  : :  :  |:    :|| :     :::::::::  : :  ::: : :
Db        499 MYMKCSYMMRYGYCKACKKCCYAMCWKAAYSGMMMYWYRKYSKWMRMSTKYMWSMWYKKC 440

Qy        409 GAGATGTTGGCCCTCAACCTGGGGGTCACCCGGCAGACCGTCCACGCCTGGAAGAAGGTC 468
              :  :| :| : |::: :|  |  ::: :: :   ::|:::: :: :      :   : ::
Db        439 RSMKYGAKGCYGCKMWTYCSYGYMKWYTYMGSYKYSRCYKYMRMYMYKGWMYMMYYSAYS 380

Qy        469 CTTGAGAAAAAGGGCCTGGTGGCCACCGACGTCCTTCACCAAACCGTCAACGGGGAGCGC 528
              :   :    |   :: ::: | :::: : ::: : | :|::  :  : ::   ::     :
Db        379 SMMTWYYYYAKYWKYWYKRRGTMSWYGKSYKKKYCTWWCYMKCMRCYRWRKMMRKKTKYS 320

Qy        529 CGGGCCATCGGCACCCTTTGGGCCGTCCGGCTGAGGCCAGGGAAAGCCAGGCTCACCCTG 588
              :  | ::   ::|||: ::| ::: :: :  |  ::   ::: ::::: ::: :  :: 
Db        319 KRCYCWRYATCYWCCCYRKRGWYSRRSMMRTAGKWKMRSWSRWCRSYSWYKMYKKMWKKS 260

Qy        589 GACGACTACATCTACCCCTGGAGGAACCTCGCCCTAGACATGGCCAACGGCGTGCTCTCC 648
              ::  :|    |:: :  ::  :  :   :  : :   ::: | : :: : :   ::|
Db        259 YYMSYGWARSSGTWSRSAAKRTYKGYSTSRRAKMMRACRMYSACRRYSRTSYYCGCSYCG 200

Qy        649 TTCAACTGGGTCAAGGCCTACCAGGACCACGGAATCCGCCCCACCCTGGACGTGCTGGTC 708
              : : ::: ::: |: :|:    |:      : || ::::: :|     |::  ::::
Db        199 SSKWKYMSKSCSMRMTCSSWCSCCYTCYYGAMCWSCCMSMMYMGSCGCYTRGWKWRSKYS 140

Qy        709 CTCTGGGCTCAGGGGAAAAGGGTGATGCCCAACACCAAGACCGTGGCCGTTGACCTGGGC 768
              : |::  : | :  :     :  :::::::  | :|   ::::: ::: :   :  ::
Db        139 MCCKKYCSCCTKYCSYTGYYRYCKWYKYSYYKCYYCYCYWYMSYMRYMMKCMCSRSCSSW 80

Qy        769 CTCATCCTGGTCCTCCCCGAGGTGGAGCGTTCCAAACTCCCGGCCCTTATCACCCTCATT 828
              : ||:|:|::|: ::::: | : |  |::  : :  :  ::::::: : |    || :::
Db         79 MSCAYCSTSSTSRWMSMYYAAKMGMCGSSGMYRMSKSCKMYSKYSSCKYTGSKKCTKRKY 20
```

RESULT 4
ADA70342
ID ADA70342 standard; DNA; 2030 BP.
XX
AC ADA70342;
XX
DT 20-NOV-2003 (first entry)
XX
DE Rice gene, SEQ ID 3665.
XX
KW Plant; bacterial infection; fungal infection; viral infection; rice;
KW gene; ds.
XX
OS Oryza sativa.
XX
PN WO2003000898-A1.
XX
PD 03-JAN-2003.
XX
PF 22-JUN-2001; 2001WO-IB001105.
XX
PR 22-JUN-2001; 2001WO-IB001105.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX
DR WPI; 2003-175290/17.
XX
PT Identifying at least one gene involved in plant resistance or response to
PT pathogenic infection for conferring resistance or tolerance to a plant to
PT bacterial, fungal or viral infection by determining or detecting plant
PT gene expression.
XX
PS Claim 6; SEQ ID NO 3665; 899pp; English.
XX

CC The present invention relates to a method (M1) for identifying genes
CC involved in plant resistance or response to pathogenic infection. M1
CC comprises identifying a gene whose expression is significantly altered in
CC the incompatible interaction of plant gene expression relative to
CC expression of the gene in an uninfected plant, in a mutant plant that
CC does not express a gene associated with response to pathogenic infection,
CC or in a corresponding incompatible or compatible interaction. (M1) is
CC useful for conferring resistance to resistance or tolerance to a plant to
CC bacterial, fungal or viral infection. The present sequence was used to
CC illustrate the invention.
XX
SQ Sequence 2030 BP; 388 A; 659 C; 727 G; 246 T; 0 U; 10 Other;

Query Match 5.5%; Score 56.4; DB 7; Length 2030;
Best Local Similarity 46.7%; Pred. No. 0.0023;
Matches 214; Conservative 0; Mismatches 241; Indels 3; Gaps 1;

```
Qy        356 GCAACCGGGAGCTCTGGAGGAAGGTGGGGACGGTCGTCTTCATGGTCCCCCTGGAGATGT 415
              ||  |  || ||||  | ||| |  || ||||| ||    |   ||  |  | || |
Db        485 GCGCCATGGTGCTCACGCGGATGAAGGAGACGGCCGAGGCCTACCTCGGCGAGAAGGTCA 544

Qy        416 TGGCCCTCAACCTGGGGGTCACCCGGCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGA 475
              |  |  |  |  ||| |      |  |  | ||||   | | ||| |    ||
Db        545 CGCGCGCCGTCGTCACCGTCCCGGCCTACTTCAACGACGCGCAGCGGCAGGCCACCAAGG 604

Qy        476 AAAAGGGCCTGGTGGCCACCGACGTCCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCA 535
              |   ||| |  | |||     |  |  |||  | | ||||| |  |  |   |||
Db        605 ACGCCGGCGTCATCGCCGGGCTCACCGTCGACCGCATCATCAACGAGCCGACCGCCGCCG 664

Qy        536 TCGGCACCCTTTGGGCCGTCCGGCTGAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACT 595
              |  | ||   |  || | |  | | | | ||   |  |||  | | ||| |
Db        665 CCATCGCCTACGGCATCGACAAGAAGGGCGCCGAGAAGAACGTCCTCGTCTTCGACCTCG 724

Qy        596 ACATCTACCCCTGGAGGAACCTCGCCCTAGACATGGCCAACGGCGTGCTCTCCTTCAACT 655
              |  |  | | |          | ||| | ||| | |||||||||| ||   ||
Db        725 GCGGCGGCACGTTCGACGTCAGCATCCTCGCCATCGACAACGGCGTGTTCGAGGTCCTTG 784

Qy        656 GGGTCAAGGCCTACCAGGACCACGGAATC---CGCCCCACCCTGGACGTGCTGGTCCTCT 712
                ||| | | ||   ||| |||   |   | |  ||   | | ||| || ||
Db        785 CCACCAACGGCGACACCCACCTCGGCGGCGAGGACTTCGACCAACGCCTCATGGACCACT 844

Qy        713 GGGCTCAGGGGAAAAGGGTGATGCCCAACACCAAGACCGTGGCCGTTGACCTGGGCCTCA 772
              |  ||| |   |  || || | |  | | || | ||  |  |  |   ||
Db        845 TCGTCAAGGTCATCCGCCGGAAGCACGGGCGCGACATCGCCGGCGACGCGCGCGCGCTGG 904

Qy        773 TCCTGGTCCTCCCCGAGGTGGAGCGTTCCAAACTCCCG 810
              |  | ||| || ||||   |||||  |||| | | ||
Db        905 GCAAGCTCCGCCGCGAGTGCGAGCGCGCCAAGCGCGCG 942
```

RESULT 5
ADC08176
ID ADC08176 standard; DNA; 2031 BP.
XX
AC ADC08176;
XX
DT 18-DEC-2003 (first entry)
XX
DE Rice DNA sequence Seq ID443 related to grain filling.
XX
KW plant biotechnology; carbohydrate synthesis; carbohydrate metabolism;
KW carbohydrate degradation; carbohydrate; plant grain; grain filling; corn;
KW tomato; banana; canola; cotton; peanut; sorghum; tobacco; sugarbeet;
KW wheat; rice; protein; oil; starch; fibre; moisture content; cereal grain;
KW gene; ds; plant.
XX
OS Oryza sativa.
XX
PN WO2003000905-A2.
XX
PD 03-JAN-2003.
XX
PF 21-JUN-2002; 2002WO-IB002450.

XX
PR 22-JUN-2001; 2001US-0300112P.
PR 26-SEP-2001; 2001US-0325277P.
PR 20-DEC-2001; 2001US-0342327P.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Zhu T, Cheng W, Briggs S, Cooper B, Goff SA, Moughamer T;
PI Glazebrook J, Katagiri F, Kreps J, Provart N, Ricke D;
XX
DR WPI; 2003-229341/22.
DR P-PSDB; ADC08177.
XX
PT New plant genes encoding polypeptides having an activity involved in or
PT associated with the synthesis, metabolism or degradation of carbohydrates
PT in the plant grain useful in generating plants having improved
PT nutritional properties.
XX
PS Claim 35; SEQ ID NO 443; 130pp; English.
XX
CC This invention, in the area of plant biotechnology, relates to novel
CC polynucleotides comprising a nucleotide sequence encoding a protein which
CC is involved in or associated with the synthesis, metabolism or
CC degradation of carbohydrates in the plant grain and the expression of
CC which is up-regulated during grain filling. The plant is selected from
CC corn, tomato, banana, canola, cotton, peanut, sorghum, tobacco,
CC sugarbeet, wheat, and rice. The invention may be useful for the
CC improvement of protein, oil, starch, fibre and moisture content of the
CC cereal grains. In addition, carbohydrate levels may be modified to a more
CC desirable level using the present invention. The present sequence is a
CC DNA sequence encoding a rice protein of the invention. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/publishedpct_sequences.
XX
SQ Sequence 2031 BP; 388 A; 659 C; 728 G; 246 T; 0 U; 10 Other;

Query Match 5.5%; Score 56.4; DB 9; Length 2031;
Best Local Similarity 46.7%; Pred. No. 0.0023;
Matches 214; Conservative 0; Mismatches 241; Indels 3; Gaps 1;

Qy 356 GCAACCGGGAGCTCTGGAGGAAGGTGGGGACGGTCGTCTTCATGGTCCCCCTGGAGATGT 415
Db 485 GCGCCATGGTGCTCACGCGGATGAAGGAGACGGCCGAGGCCTACCTCGGCGAGAAGGTCA 544

Qy 416 TGGCCCTCAACCTGGGGGTCACCCGGCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGA 475
Db 545 CGCGCGCCGTCGTCACCGTCCCGGCCTACTTCAACGACGCGCAGCGGCAGGCCACCAAGG 604

Qy 476 AAAAGGGCCTGGTGGCCACCGACGTCCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCA 535
Db 605 ACGCCGGCGTCATCGCCGGGCTCACCGTCGACCGCATCATCAACGAGCCGACCGCCGCCG 664

Qy 536 TCGGCACCCTTTGGGCCGTCCGGCTGAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACT 595
Db 665 CCATCGCCTACGGCATCGACAAGAAGGGCGCCGAGAAGAACGTCCTCGTCTTCGACCTCG 724

Qy 596 ACATCTACCCCTGGAGGAACCTCGCCCTAGACATGGCCAACGGCGTGCTCTCCTTCAACT 655
Db 725 GCGGCGGCACGTTCGACGTCAGCATCCTCGCCATCGACAACGGCGTGTTCGAGGTCCTTG 784

Qy 656 GGGTCAAGGCCTACCAGGACCACGGAATC---CGCCCCACCCTGGACGTGCTGGTCCTCT 712
Db 785 CCACCAACGGCGACACCCACCTCGGCGGCGAGGACTTCGACCAACGCCTCATGGACCACT 844

Qy 713 GGGCTCAGGGGAAAAGGGTGATGCCCAACACCAAGACCGTGGCCGTTGACCTGGGCCTCA 772
Db 845 TCGTCAAGGTCATCCGCCGGAAGCACGGGCGCGACATCGCCGGCGACGCGCGCGCGCTGG 904

Qy 773 TCCTGGTCCTCCCCGAGGTGGAGCGTTCCAAACTCCCG 810
Db 905 GCAAGCTCCGCCGCGAGTGCGAGCGCGCCAAGCGCGCG 942

RESULT 6
ADA48615
ID ADA48615 standard; DNA; 2028 BP.
XX
AC ADA48615;
XX
DT 20-NOV-2003 (first entry)
XX
DE Rice gene conferring disease resistance in plants.
XX
KW disease resistance; pathogen tolerance; plant pathogen; ds; gene; plant.
XX
OS Oryza sativa.
XX
PN WO2003000906-A2.
XX
PD 03-JAN-2003.
XX
PF 21-JUN-2002; 2002WO-IB002453.
XX
PR 22-JUN-2001; 2001US-0300112P.
PR 26-SEP-2001; 2001US-0352277P.
PR 22-MAR-2002; 2002US-0366535P.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Glazebrook J, Briggs S, Cooper B, Goff SA, Moughamer T;
PI Katagiri F, Kreps J, Provart N, Ricke D, Zhu T;
XX
DR WPI; 2003-184052/18.
DR P-PSDB; ADA48616.
XX
PT New polynucleotide comprising a plant nucleotide sequence having an open
PT reading frame that encodes a polypeptide associated with disease
PT resistance, useful for conferring resistance or tolerance to a plant
PT pathogen.
XX
PS Claim 1; SEQ ID NO 685; 299pp; English.
XX
CC The invention relates to a novel isolated polynucleotide comprising a
CC plant nucleotide sequence having an open reading frame that encodes a
CC polypeptide associated with disease resistance or its fragment having
CC substantially the same activity as the full-length polypeptide. The
CC polynucleotide of the invention is useful for conferring resistance or
CC tolerance to a plant pathogen. The present sequence represents a gene
CC conferring disease resistance used in the invention.
XX
SQ Sequence 2028 BP; 390 A; 662 C; 729 G; 246 T; 0 U; 1 Other;

Query Match 5.4%; Score 55.4; DB 8; Length 2028;
Best Local Similarity 46.5%; Pred. No. 0.004;
Matches 213; Conservative 0; Mismatches 242; Indels 3; Gaps 1;

Qy 356 GCAACCGGGAGCTCTGGAGGAAGGTGGGGACGGTCGTCTTCATGGTCCCCCTGGAGATGT 415
Db 485 GCGCCATGGTGCTCACCCGGATGAAGGAGACGGCCGAGGCCTACCTCGGCGAGAAGGTCA 544

Qy 416 TGGCCCTCAACCTGGGGGTCACCCGGCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGA 475
Db 545 CGCGCGCCGTCGTCACCGTCCCGGCCTACTTCAACGACGCGCAGCGGCAGGCCACCAAGG 604

Qy 476 AAAAGGGCCTGGTGGCCACCGACGTCCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCA 535
Db 605 ACGCCGGCGTCATCGCCGGGCTCACCGTCGACCGCATCATCAACGAGCCGACCGCCGCCG 664

Qy 536 TCGGCACCCTTTGGGCCGTCCGGCTGAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACT 595
Db 665 CCATCGCCTACGGCATCGACAAGAAGGGCGCNGAGAAGAACGTCCTCGTCTTCGACCTCG 724

Qy 596 ACATCTACCCCTGGAGGAACCTCGCCCTAGACATGGCCAACGGCGTGCTCTCCTTCAACT 655
Db 725 GCGGCGGCACGTTCGACGTCAGCATCCTCGCCATCGACAACGGCGTGTTCGAGGTCCTTG 784

```
Qy        656 GGGTCAAGGCCTACCAGGACCACGGAATC---CGCCCCACCCTGGACGTGCTGGTCCTCT 712
              ||| | | ||    ||| |||   |     | |  ||    | |  ||| || ||
Db        785 CCACCAACGGCGACACCCACCTCGGCGGCGAGGACTTCGACCAACGCCTCATGGACCACT 844

Qy        713 GGGCTCAGGGGAAAAGGGTGATGCCCAACACCAAGACCGTGGCCGTTGACCTGGGCCTCA 772
              |   |||  |   |  || || |    | | ||  | ||  |  |  |  ||
Db        845 TCGTCAAGGTCATCCGCCGGAAGCACGGGCGCGACATCGCCGGCGACGCGCGCGCGCTGG 904

Qy        773 TCCTGGTCCTCCCCGAGGTGGAGCGTTCCAAACTCCCG 810
              |  | ||| || ||||   ||||| |||| | | ||
Db        905 GCAAGCTCCGCCGCGAGTGCGAGCGCGCCAAGCGCGCG 942
```

RESULT 7
ADC08212
ID ADC08212 standard; DNA; 2028 BP.
XX
AC ADC08212;
XX
DT 18-DEC-2003 (first entry)
XX
DE Rice DNA sequence Seq ID517 related to grain filling.
XX
KW plant biotechnology; carbohydrate synthesis; carbohydrate metabolism;
KW carbohydrate degradation; carbohydrate; plant grain; grain filling; corn;
KW tomato; banana; canola; cotton; peanut; sorghum; tobacco; sugarbeet;
KW wheat; rice; protein; oil; starch; fibre; moisture content; cereal grain;
KW gene; ds; plant.
XX
OS Oryza sativa.
XX
PN WO2003000905-A2.
XX
PD 03-JAN-2003.
XX
PF 21-JUN-2002; 2002WO-IB002450.
XX
PR 22-JUN-2001; 2001US-0300112P.
PR 26-SEP-2001; 2001US-0325277P.
PR 20-DEC-2001; 2001US-0342327P.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Zhu T, Cheng W, Briggs S, Cooper B, Goff SA, Moughamer T;
PI Glazebrook J, Katagiri F, Kreps J, Provart N, Ricke D;
XX
DR WPI; 2003-229341/22.
DR P-PSDB; ADC08213.
XX
PT New plant genes encoding polypeptides having an activity involved in or
PT associated with the synthesis, metabolism or degradation of carbohydrates
PT in the plant grain useful in generating plants having improved
PT nutritional properties.
XX
PS Claim 35; SEQ ID NO 517; 130pp; English.
XX
CC This invention, in the area of plant biotechnology, relates to novel
CC polynucleotides comprising a nucleotide sequence encoding a protein which
CC is involved in or associated with the synthesis, metabolism or
CC degradation of carbohydrates in the plant grain and the expression of
CC which is up-regulated during grain filling. The plant is selected from
CC corn, tomato, banana, canola, cotton, peanut, sorghum, tobacco,
CC sugarbeet, wheat, and rice. The invention may be useful for the
CC improvement of protein, oil, starch, fibre and moisture content of the
CC cereal grains. In addition, carbohydrate levels may be modified to a more
CC desirable level using the present invention. The present sequence is a
CC DNA sequence encoding a rice protein of the invention. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/publishedpct_sequences.
XX
SQ Sequence 2028 BP; 390 A; 662 C; 729 G; 246 T; 0 U; 1 Other;

Query Match 5.4%; Score 55.4; DB 9; Length 2028;
Best Local Similarity 46.5%; Pred. No. 0.004;
Matches 213; Conservative 0; Mismatches 242; Indels 3; Gaps 1;

```
Qy        356 GCAACCGGGAGCTCTGGAGGAAGGTGGGGACGGTCGTCTTCATGGTCCCCCTGGAGATGT 415
              ||  |  ||  ||||    |||  |  ||  ||||| ||     |    ||  |  | || |
Db        485 GCGCCATGGTGCTCACCCGGATGAAGGAGACGGCCGAGGCCTACCTCGGCGAGAAGGTCA 544

Qy        416 TGGCCCTCAACCTGGGGGTCACCCGGCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGA 475
              |  |  |  |  |    ||| |     |  |  ||||    |  |  |||  |    ||
Db        545 CGCGCGCCGTCGTCACCGTCCCGGCCTACTTCAACGACGCGCAGCGGCAGGCCACCAAGG 604

Qy        476 AAAAGGGCCTGGTGGCCACCGACGTCCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCA 535
              |    |||  |  |||     |  |  |  |||  |  | ||||||  |  |  |  |||
Db        605 ACGCCGGCGTCATCGCCGGGCTCACCGTCGACCGCATCATCAACGAGCCGACCGCCGCCG 664

Qy        536 TCGGCACCCTTTGGGCCGTCCGGCTGAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACT 595
              |  | ||  |  || |  |  |  |  |  | |||  |  |||  |  | |||  |
Db        665 CCATCGCCTACGGCATCGACAAGAAGGGCGCNGAGAAGAACGTCCTCGTCTTCGACCTCG 724

Qy        596 ACATCTACCCCTGGAGGAACCTCGCCCTAGACATGGCCAACGGCGTGCTCTCCTTCAACT 655
              |  |  |  |  |     |  |  ||| | ||| | ||||||||||  ||    ||
Db        725 GCGGCGGCACGTTCGACGTCAGCATCCTCGCCATCGACAACGGCGTGTTCGAGGTCCTTG 784

Qy        656 GGGTCAAGGCCTACCAGGACCACGGAATC---CGCCCCACCCTGGACGTGCTGGTCCTCT 712
              |||  | | ||    |||  |||   |     | |  ||    | |  ||| || ||
Db        785 CCACCAACGGCGACACCCACCTCGGCGGCGAGGACTTCGACCAACGCCTCATGGACCACT 844

Qy        713 GGGCTCAGGGGAAAAGGGTGATGCCCAACACCAAGACCGTGGCCGTTGACCTGGGCCTCA 772
              |   |||  |   |  || || |    | | ||  | ||  |  |  |  ||
Db        845 TCGTCAAGGTCATCCGCCGGAAGCACGGGCGCGACATCGCCGGCGACGCGCGCGCGCTGG 904

Qy        773 TCCTGGTCCTCCCCGAGGTGGAGCGTTCCAAACTCCCG 810
              |  | ||| || ||||   ||||| |||| | | ||
Db        905 GCAAGCTCCGCCGCGAGTGCGAGCGCGCCAAGCGCGCG 942
```

RESULT 8
ADA70341
ID ADA70341 standard; DNA; 2028 BP.
XX
AC ADA70341;
XX
DT 20-NOV-2003 (first entry)
XX
DE Rice gene, SEQ ID 3664.
XX
KW Plant; bacterial infection; fungal infection; viral infection; rice;
KW gene; ds.
XX
OS Oryza sativa.
XX
PN WO2003000898-A1.
XX
PD 03-JAN-2003.
XX
PF 22-JUN-2001; 2001WO-IB001105.
XX
PR 22-JUN-2001; 2001WO-IB001105.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX
DR WPI; 2003-175290/17.
XX
PT Identifying at least one gene involved in plant resistance or response to
PT pathogenic infection for conferring resistance or tolerance to a plant to
PT bacterial, fungal or viral infection by determining or detecting plant
PT gene expression.
XX

PS Claim 6; SEQ ID NO 3664; 899pp; English.
XX
CC The present invention relates to a method (M1) for identifying genes
CC involved in plant resistance or response to pathogenic infection. M1
CC comprises identifying a gene whose expression is significantly altered in
CC the incompatible interaction of plant gene expression relative to
CC expression of the gene in an uninfected plant, in a mutant plant that
CC does not express a gene associated with response to pathogenic infection,
CC or in a corresponding incompatible or compatible interaction. (M1) is
CC useful for conferring resistance to resistance or tolerance to a plant to
CC bacterial, fungal or viral infection. The present sequence was used to
CC illustrate the invention.
XX
SQ Sequence 2028 BP; 390 A; 662 C; 729 G; 246 T; 0 U; 1 Other;

Query Match 5.3%; Score 54.8; DB 7; Length 2028;
Best Local Similarity 46.5%; Pred. No. 0.0056;
Matches 213; Conservative 0; Mismatches 242; Indels 3; Gaps 1;

Qy 356 GCAACCGGGAGCTCTGGAGGAAGGTGGGGACGGTCGTCTTCATGGTCCCCCTGGAGATGT 415
Db 485 GCGCCATGGTGCTCACCCGGATGAAGGAGACGGCCGAGGCCTACCTCGGCGAGAAGGTCA 544

Qy 416 TGGCCCTCAACCTGGGGGTCACCCGGCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGA 475
Db 545 CGCGCGCCGTCGTCACCGTCCCGGCCTACTTCAACGACGCGCAGCGGCAGGCCACCAAGG 604

Qy 476 AAAAGGGCCTGGTGGCCACCGACGTCCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCA 535
Db 605 ACGCCGGCGTCATCGCCGGGCTCACCGTCGACCGCATCATCAACGAGCCGACCGCCGCCG 664

Qy 536 TCGGCACCCTTTGGGCCGTCCGGCTGAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACT 595
Db 665 CCATCGCCTACGGCATCGACAAGAAGGGCGCSGAGAAGAACGTCCTCGTCTTCGACCTCG 724

Qy 596 ACATCTACCCCTGGAGGAACCTCGCCCTAGACATGGCCAACGGCGTGCTCTCCTTCAACT 655
Db 725 GCGGCGGCACGTTCGACGTCAGCATCCTCGCCATCGACAACGGCGTGTTCGAGGTCCTTG 784

Qy 656 GGGTCAAGGCCTACCAGGACCACGGAATCCG---CCCCACCCTGGACGTGCTGGTCCTCT 712
Db 785 CCACCAACGGCGACACCCACCTCGGCGGCGAGGACTTCGACCAACGCCTCATGGACCACT 844

Qy 713 GGGCTCAGGGGAAAAGGGTGATGCCCAACACCAAGACCGTGGCCGTTGACCTGGGCCTCA 772
Db 845 TCGTCAAGGTCATCCGCCGGAAGCACGGGCGCGACATCGCCGGCGACGCGCGCGCGCTGG 904

Qy 773 TCCTGGTCCTCCCCGAGGTGGAGCGTTCCAAACTCCCG 810
Db 905 GCAAGCTCCGCCGCGAGTGCGAGCGCGCCAAGCGCGCG 942

RESULT 9
ADA71232
ID ADA71232 standard; DNA; 2060 BP.
XX
AC ADA71232;
XX
DT 20-NOV-2003 (first entry)
XX
DE Rice gene, SEQ ID 4555.
XX
KW Plant; bacterial infection; fungal infection; viral infection; rice;
KW gene; ds.
XX
OS Oryza sativa.
XX
PN WO2003000898-A1.
XX
PD 03-JAN-2003.
XX
PF 22-JUN-2001; 2001WO-IB001105.
XX
PR 22-JUN-2001; 2001WO-IB001105.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX
DR WPI; 2003-175290/17.
XX
PT Identifying at least one gene involved in plant resistance or response to
PT pathogenic infection for conferring resistance or tolerance to a plant to
PT bacterial, fungal or viral infection by determining or detecting plant
PT gene expression.
XX
PS Claim 6; SEQ ID NO 4555; 899pp; English.
XX
CC The present invention relates to a method (M1) for identifying genes
CC involved in plant resistance or response to pathogenic infection. M1
CC comprises identifying a gene whose expression is significantly altered in
CC the incompatible interaction of plant gene expression relative to
CC expression of the gene in an uninfected plant, in a mutant plant that
CC does not express a gene associated with response to pathogenic infection,
CC or in a corresponding incompatible or compatible interaction. (M1) is
CC useful for conferring resistance to resistance or tolerance to a plant to
CC bacterial, fungal or viral infection. The present sequence was used to
CC illustrate the invention.
XX
SQ Sequence 2060 BP; 395 A; 670 C; 739 G; 254 T; 0 U; 2 Other;

Query Match 5.3%; Score 54.8; DB 7; Length 2060;
Best Local Similarity 46.5%; Pred. No. 0.0056;
Matches 213; Conservative 0; Mismatches 242; Indels 3; Gaps 1;

Qy 356 GCAACCGGGAGCTCTGGAGGAAGGTGGGGACGGTCGTCTTCATGGTCCCCCTGGAGATGT 415
Db 485 GCGCCATGGTGCTCACCCGGATGAAGGAGACGGCCGAGGCCTACCTCGGCGAGAAGGTCA 544

Qy 416 TGGCCCTCAACCTGGGGGTCACCCGGCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGA 475
Db 545 CGCGCGCCGTCGTCACCGTCCCGGCCTACTTCAACGACGCGCAGCGGCAGGCCACCAAGG 604

Qy 476 AAAAGGGCCTGGTGGCCACCGACGTCCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCA 535
Db 605 ACGCCGGCGTCATCGCCGGGCTCACCGTCGACCGCATCATCAACGAGCCGACCGCCGCCG 664

Qy 536 TCGGCACCCTTTGGGCCGTCCGGCTGAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACT 595
Db 665 CCATCGCCTACGGCATCGACAAGAAGGGCGCGGAGAAGAACGTCCTCGTCTTCGACCTCG 724

Qy 596 ACATCTACCCCTGGAGGAACCTCGCCCTAGACATGGCCAACGGCGTGCTCTCCTTCAACT 655
Db 725 GCGGCGGCACGTTCGACGTCAGCATCCTCGCCATCGACAACGGCGTGTTCGAGGTCCTTG 784

Qy 656 GGGTCAAGGCCTACCAGGACCACGGAATCCG---CCCCACCCTGGACGTGCTGGTCCTCT 712
Db 785 CCACCAACGGCGACACCCACCTCGGCGGCGAGGACTTCGACCAACGCCTCATGGACCACT 844

Qy 713 GGGCTCAGGGGAAAAGGGTGATGCCCAACACCAAGACCGTGGCCGTTGACCTGGGCCTCA 772
Db 845 TCGTCAAGGTCATCCGCCGGAAGCACGGGCGCGACATCGCCGGCGACGCGCGCGCGCTGG 904

Qy 773 TCCTGGTCCTCCCCGAGGTGGAGCGTTCCAAACTCCCG 810
Db 905 GCAAGCTCCGCCGCGAGTGCGAGCGCGCCAAGCGCGCG 942

RESULT 10
AAQ37036
ID AAQ37036 standard; DNA; 615 BP.
XX
AC AAQ37036;
XX
DT 05-JUL-1993 (first entry)

XX
DE SOD-T gene.
XX
KW Super oxide dismutase; heat resistant; medicine; inflammation; X-ray;
KW enzyme electrodes; bioreactor; ss.
XX
OS Thermus aquaticus.
XX
PN JP05030976-A.
XX
PD 09-FEB-1993.
XX
PF 26-JUL-1991; 91JP-00208747.
XX
PR 26-JUL-1991; 91JP-00208747.
XX
PA (YOTS-) YOTSUBA NYUGYO KK.
XX
DR WPI; 1993-087958/11.
DR P-PSDB; AAR32376.
XX
PT Gene, of and plasmid contg. heat resistant super:oxide dismutase - for
PT treating inflammation, X=ray protection, enzyme electrodes and
PT bio:reactor enzymes.
XX
PS Claim 1; Page 7; 9pp; Japanese.
XX
CC The sequence shown is that isolated from Thermus aquaticus genomic DNA
CC and encodes a heat resistant super oxide dismutase enzyme. The gene is
CC useful for medicines for inflammation and as a protection against X-rays,
CC as a reagent for research in enzyme electrodes or as an enzyme for a
CC bioreactor
XX
SQ Sequence 615 BP; 125 A; 223 C; 183 G; 84 T; 0 U; 0 Other;

Query Match 5.3%; Score 54.4; DB 2; Length 615;
Best Local Similarity 47.5%; Pred. No. 0.0051;
Matches 194; Conservative 0; Mismatches 211; Indels 3; Gaps 1;

```
Qy        336 CCGCTCCACCATCCCCTACGGCAACCGGGAGCTCTGGAGGAAGGTGGGGACGGTCGTCTT 395
Db        186 CCTCACCGCCCTCCCCGCCGACATCCAGGCCGCCCTGCGCAACAACGGGGGCGGGCACCT 245

Qy        396 CATGGTCCCCCTGGAGATGTTGGCCCTCAACCTGGGGGTCACCCGGCAGACCGTCCACGC 455
Db        246 GAAGCACAGCCTCTTCTGGCGCCTCCTCACCCCGGGCGGGGCCAAGGAGCCCGTGGGGGA 305

Qy        456 CTGGAAGAAGGTCCTTGAGAAAAAGGGCCTGGTGGCCACCGACGTCCTTCACCAAACCGT 515
Db        306 GCTGAAGAAGGCCATTGACGAGCAGTTTGGCGGCTTCGCCGCCCTAAAGGAGAAGCTCAC 365

Qy        516 CAACGGGGAGCGCCGGGCCATCGGCACCCTTTGGGCCGTCCGGCTGAGGCCAGGGAAAGC 575
Db        366 CCAGGCGGCCATGCGGCGCTTCGGCTCCGGGTGGGCCTGGCTGGTCAAGGAGCCCTTCGG 425

Qy        576 CAGGCTCACCCTGGACGACTACATCTACCCCTGGAGGAACCTCGCCCTAGACATGGCCAA 635
Db        426 CAAGCTCCACGTGATCTCCACCGCCAACCAGGACAACCCAGTGATGGGAGGCTTCGCCCC 485

Qy        636 CGGCGTGCTCTCCTTCAACTGGGTCAAGGC---CTACCAGGACCACGGAATCCGCCCCAC 692
Db        486 CATCGTGGGCATTGACGTCTGGGAGCACGCCTACTACCTGAAGTACCAGAACGGCCGGGC 545

Qy        693 CCTGGACGTGCTGGTCCTCTGGGCTCAGGGGAAAAGGGTGATGCCCAA 740
Db        546 CGACTACCTGCAGGCCATCTGGAACGTCCTGAACTGGGAGGTGGCCGA 593
```

RESULT 11
ADA70624
ID ADA70624 standard; DNA; 1755 BP.
XX
AC ADA70624;
XX
DT 20-NOV-2003 (first entry)
XX
DE Rice gene, SEQ ID 3947.
XX
KW Plant; bacterial infection; fungal infection; viral infection; rice;
KW gene; ds.
XX
OS Oryza sativa.
XX
PN WO2003000898-A1.
XX
PD 03-JAN-2003.
XX
PF 22-JUN-2001; 2001WO-IB001105.
XX
PR 22-JUN-2001; 2001WO-IB001105.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX
DR WPI; 2003-175290/17.
XX
PT Identifying at least one gene involved in plant resistance or response to
PT pathogenic infection for conferring resistance or tolerance to a plant to
PT bacterial, fungal or viral infection by determining or detecting plant
PT gene expression.
XX
PS Claim 6; SEQ ID NO 3947; 899pp; English.
XX
CC The present invention relates to a method (M1) for identifying genes
CC involved in plant resistance or response to pathogenic infection. M1
CC comprises identifying a gene whose expression is significantly altered in
CC the incompatible interaction of plant gene expression relative to
CC expression of the gene in an uninfected plant, in a mutant plant that
CC does not express a gene associated with response to pathogenic infection,
CC or in a corresponding incompatible or compatible interaction. (M1) is
CC useful for conferring resistance to resistance or tolerance to a plant to
CC bacterial, fungal or viral infection. The present sequence was used to
CC illustrate the invention.
XX
SQ Sequence 1755 BP; 341 A; 592 C; 527 G; 293 T; 0 U; 2 Other;

Query Match 5.2%; Score 53.2; DB 7; Length 1755;
Best Local Similarity 50.8%; Pred. No. 0.013;
Matches 127; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

```
Qy        543 CCTTTGGGCCGTCCGGCTGAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACTACATCTA 602
Db        246 CGTCTTCGATGGCCGGGAGCTCGCCCGGAACGACAGCCGCATCATGGACTACGGCGTCTC 305

Qy        603 CCCCTGGAGGAACCTCGCCCTAGACATGGCCAACGGCGTGCTCTCCTTCAACTGGGTCAA 662
Db        306 CCACGGCAACGTCCTCCACCTCGTCATCCGCATCTCCGACCTCCGCCTCATCACCGTCCA 365

Qy        663 GGCCTACCAGGACCACGGAATCCGCCCCACCCTGGACGTGCTGGTCCTCTGGGCTCAGGG 722
Db        366 GACCGTCCACGGCAACAAGTTCAGGTTCCGCGTCGAGCCCGGCCGCACCGTCGGCTACGT 425

Qy        723 GAAAAGGGTGATGCCCAACACCAAGACCGTGGCCGTTGACCTGGGCCTCATCCTGGTCCT 782
Db        426 CAAGCAGCAGATCGCCAAGAACTCCACCCACGACGACGACCACCACAGCCTCGTCCTCCA 485

Qy        783 CCCCGAGGTG 792
Db        486 GGGCGAGGTG 495
```

RESULT 12
ADA48223
ID ADA48223 standard; DNA; 1965 BP.

XX
AC ADA48223;
XX
DT 20-NOV-2003 (first entry)
XX
DE Rice gene conferring disease resistance in plants.
XX
KW disease resistance; pathogen tolerance; plant pathogen; ds; gene; plant.
XX
OS Oryza sativa.
XX
PN WO2003000906-A2.
XX
PD 03-JAN-2003.
XX
PF 21-JUN-2002; 2002WO-IB002453.
XX
PR 22-JUN-2001; 2001US-0300112P.
PR 26-SEP-2001; 2001US-0352277P.
PR 22-MAR-2002; 2002US-0366535P.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Glazebrook J, Briggs S, Cooper B, Goff SA, Moughamer T;
PI Katagiri F, Kreps J, Provart N, Ricke D, Zhu T;
XX
DR WPI; 2003-184052/18.
DR P-PSDB; ADA48224.
XX
PT New polynucleotide comprising a plant nucleotide sequence having an open
PT reading frame that encodes a polypeptide associated with disease
PT resistance, useful for conferring resistance or tolerance to a plant
PT pathogen.
XX
PS Claim 1; SEQ ID NO 293; 299pp; English.
XX
CC The invention relates to a novel isolated polynucleotide comprising a
CC plant nucleotide sequence having an open reading frame that encodes a
CC polypeptide associated with disease resistance or its fragment having
CC substantially the same activity as the full-length polypeptide. The
CC polynucleotide of the invention is useful for conferring resistance or
CC tolerance to a plant pathogen. The present sequence represents a protein
CC conferring disease resistance used in the invention.
XX
SQ Sequence 1965 BP; 383 A; 646 C; 699 G; 235 T; 0 U; 2 Other;

Query Match 5.2%; Score 53.2; DB 8; Length 1965;
Best Local Similarity 46.3%; Pred. No. 0.013;
Matches 212; Conservative 0; Mismatches 243; Indels 3; Gaps 1;

```
Qy        356 GCAACCGGGAGCTCTGGAGGAAGGTGGGGACGGTCGTCTTCATGGTCCCCCTGGAGATGT 415
              ||  |  || ||||  | |||| |  || ||||| ||    |    ||  |  | || |
Db        437 GCGCCATGGTGCTCACGCGGATGAAGGAGACGGCCGAGGCCTACCTCGGCGAGAAGGTCA 496

Qy        416 TGGCCCTCAACCTGGGGGTCACCCGGCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGA 475
              |  |  | |    ||| |      |   | | ||||   |  | ||| |    ||
Db        497 CCCGCGCCGTCGTCACCGTCCCGGCCTACTTCAACGACGCGCAGCGGCAGGCCACCAAGG 556

Qy        476 AAAAGGGCCTGGTGGCCACCGACGTCCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCA 535
              |    ||| |  | |||     | | |  ||| | | ||||||| |     |   |||
Db        557 ACGCCGGCGTCATCGCCGGGCTCACCGTCGACCGCATCATCAACGAGCCCACCGCCGCCG 616

Qy        536 TCGGCACCCTTTGGGCCGTCCGGCTGAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACT 595
              | | ||   |  || |  | | |  | | |||   |   |||  | | |||  |
Db        617 CCATCGCCTACGGCATCGACAAGAAGGGCGCCGAGAAGAACGTCCTCGTCTTCGACCTCG 676

Qy        596 ACATCTACCCCTGGAGGAACCTCGCCCTAGACATGGCCAACGGCGTGCTCTCCTTCAACT 655
              |  |  | | |      |  | ||| | ||| | ||||||||| ||    ||
Db        677 GCGGCGGCACGTTCGACGTCAGCATCCTCGCCATCGACAACGGCGTGTTCGAGGTCCTTG 736

Qy        656 GGGTCAAGGCCTACCAGGACCACGGAATCCG---CCCCACCCTGGACGTGCTGGTCCTCT 712
              |||  | | ||    ||| |||    |      |  |  ||    | |  ||| || ||
Db        737 CCACCAACGGCGACACCCACCTCGGCGGCGAGGACTTCGACCAACGCCTCATGGACCACT 796

Qy        713 GGGCTCAGGGGAAAAGGGTGATGCCCAACACCAAGACCGTGGCCGTTGACCTGGGCCTCA 772
              |   |||  |   |  || || |  |  | | | |   | ||  |  |  |  ||
Db        797 TCGTCAAGGTCATCCGCCGGAAGCACGGCCGCGACATCACCGGCGACGCGCGCGCGCTGG 856

Qy        773 TCCTGGTCCTCCCCGAGGTGGAGCGTTCCAAACTCCCG 810
              |  | ||| || ||||   |||||  |||| | | ||
Db        857 GCAAGCTCCGCCGCGAGTGCGAGCGCGCCAAGCGCGCG 894
```

RESULT 13
ADC08174
ID ADC08174 standard; DNA; 1965 BP.
XX
AC ADC08174;
XX
DT 18-DEC-2003 (first entry)
XX
DE Rice DNA sequence Seq ID441 related to grain filling.
XX
KW plant biotechnology; carbohydrate synthesis; carbohydrate metabolism;
KW carbohydrate degradation; carbohydrate; plant grain; grain filling; corn;
KW tomato; banana; canola; cotton; peanut; sorghum; tobacco; sugarbeet;
KW wheat; rice; protein; oil; starch; fibre; moisture content; cereal grain;
KW gene; ds; plant.
XX
OS Oryza sativa.
XX
PN WO2003000905-A2.
XX
PD 03-JAN-2003.
XX
PF 21-JUN-2002; 2002WO-IB002450.
XX
PR 22-JUN-2001; 2001US-0300112P.
PR 26-SEP-2001; 2001US-0325277P.
PR 20-DEC-2001; 2001US-0342327P.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Zhu T, Cheng W, Briggs S, Cooper B, Goff SA, Moughamer T;
PI Glazebrook J, Katagiri F, Kreps J, Provart N, Ricke D;
XX
DR WPI; 2003-229341/22.
DR P-PSDB; ADC08175.
XX
PT New plant genes encoding polypeptides having an activity involved in or
PT associated with the synthesis, metabolism or degradation of carbohydrates
PT in the plant grain useful in generating plants having improved
PT nutritional properties.
XX
PS Claim 35; SEQ ID NO 441; 130pp; English.
XX
CC This invention, in the area of plant biotechnology, relates to novel
CC polynucleotides comprising a nucleotide sequence encoding a protein which
CC is involved in or associated with the synthesis, metabolism or
CC degradation of carbohydrates in the plant grain and the expression of
CC which is up-regulated during grain filling. The plant is selected from
CC corn, tomato, banana, canola, cotton, peanut, sorghum, tobacco,
CC sugarbeet, wheat, and rice. The invention may be useful for the
CC improvement of protein, oil, starch, fibre and moisture content of the
CC cereal grains. In addition, carbohydrate levels may be modified to a more
CC desirable level using the present invention. The present sequence is a
CC DNA sequence encoding a rice protein of the invention. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/publishedpct_sequences.
XX
SQ Sequence 1965 BP; 383 A; 646 C; 699 G; 235 T; 0 U; 2 Other;

Query Match 5.2%; Score 53.2; DB 9; Length 1965;
Best Local Similarity 46.3%; Pred. No. 0.013;
Matches 212; Conservative 0; Mismatches 243; Indels 3; Gaps 1;

```
Qy        356 GCAACCGGGAGCTCTGGAGGAAGGTGGGGACGGTCGTCTTCATGGTCCCCCTGGAGATGT 415
Db        437 GCGCCATGGTGCTCACGCGGATGAAGGAGACGGCCGAGGCCTACCTCGGCGAGAAGGTCA 496

Qy        416 TGGCCCTCAACCTGGGGGTCACCCGGCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGA 475
Db        497 CCCGCGCCGTCGTCACCGTCCCGGCCTACTTCAACGACGCGCAGCGGCAGGCCACCAAGG 556

Qy        476 AAAAGGGCCTGGTGGCCACCGACGTCCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCA 535
Db        557 ACGCCGGCGTCATCGCCGGGCTCACCGTCGACCGCATCATCAACGAGCCCACCGCCGCCG 616

Qy        536 TCGGCACCCTTTGGGCCGTCCGGCTGAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACT 595
Db        617 CCATCGCCTACGGCATCGACAAGAAGGGCGCCGAGAAGAACGTCCTCGTCTTCGACCTCG 676

Qy        596 ACATCTACCCCTGGAGGAACCTCGCCCTAGACATGGCCAACGGCGTGCTCTCCTTCAACT 655
Db        677 GCGGCGGCACGTTCGACGTCAGCATCCTCGCCATCGACAACGGCGTGTTCGAGGTCCTTG 736

Qy        656 GGGTCAAGGCCTACCAGGACCACGGAATCCG---CCCCACCCTGGACGTGCTGGTCCTCT 712
Db        737 CCACCAACGGCGACACCCACCTCGGCGGCGAGGACTTCGACCAACGCCTCATGGACCACT 796

Qy        713 GGGCTCAGGGGAAAAGGGTGATGCCCAACACCAAGACCGTGGCCGTTGACCTGGGCCTCA 772
Db        797 TCGTCAAGGTCATCCGCCGGAAGCACGGCCGCGACATCACCGGCGACGCGCGCGCGCTGG 856

Qy        773 TCCTGGTCCTCCCCGAGGTGGAGCGTTCCAAACTCCCG 810
Db        857 GCAAGCTCCGCCGCGAGTGCGAGCGCGCCAAGCGCGCG 894
```

```
RESULT 14
AAD54645
ID   AAD54645 standard; DNA; 113193 BP.
XX
AC   AAD54645;
XX
DT   26-JUN-2003  (first entry)
XX
DE   Streptomyces nodosus amphotericin (amph) biosynthetic gene cluster.
XX
KW   Polyene; antibiotic; amphotericin; amph; polyketide; enzyme; gene; ds.
XX
OS   Streptomyces nodosus.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement(4. .1824)
FT                   /*tag= a
FT                   /product= "ABC transporter encoded by S. nodosus amphG
FT                   gene"
FT   CDS             complement(1805. .3628)
FT                   /*tag= b
FT                   /product= "ABC transporter encoded by S. nodosus amphH
FT                   gene"
FT   CDS             3840. .4874
FT                   /*tag= c
FT                   /product= "GDP-mannose dehydratase encoded by S. nodosus
FT                   amphDIII gene"
FT   CDS             5042. .33574
FT                   /*tag= d
FT                   /product= "Polyketide synthase multienzyme housing
FT                   extension modules 9, 10, 11, 12, 13 and 14 encoded by S.
FT                   nodosus amphI gene"
FT   CDS             33584. .50518
FT                   /*tag= e
FT                   /product= "Polyketide synthase multienzyme housing
FT                   extension modules 15, 16 and 17 encoded by S. nodosus
FT                   amphJ gene"
FT   CDS             50571. .56675
FT                   /*tag= f
FT                   /product= "Polyketide synthase multienzyme housing
FT                   extension modules 18 and thioesterase encoded by S.
FT                   nodosus amphK gene"
FT   CDS             56829. .58019
FT                   /*tag= g
FT                   /product= "Cytochrome P450 encoded by S. nodosus amphL
FT                   gene"
FT   CDS             58139. .58648
FT                   /*tag= h
FT                   /product= "ORF1, hypothetical protein"
FT   CDS             complement(58756. .59610)
FT                   /*tag= i
FT                   /product= "ORF2, hypothetical protein"
FT   CDS             59869. .61470
FT                   /*tag= j
FT                   /product= "ORF3, hypothetical protein"
FT                   /transl_except= (pos:59869. .59871, aa:Met)
FT   CDS             complement(61798. .61995)
FT                   /*tag= k
FT                   /product= "Ferredoxin encoded by S. nodosus amphM gene"
FT   CDS             complement(62051. .63250)
FT                   /*tag= l
FT                   /product= "Cytochrome P450 encoded by S. nodosus amphN
FT                   gene"
FT   CDS             complement(63250. .64308)
FT                   /*tag= m
FT                   /product= "NDP-sugar aminotransferase encoded by S.
FT                   nodosus amphDII gene"
FT   CDS             complement(64324. .65775)
FT                   /*tag= n
FT                   /product= "Glycosyl transferase encoded by S. nodosus
FT                   amphDI gene"
FT                   /transl_except= (pos:65773. .65775, aa:Met)
FT   CDS             66081. .70319
FT                   /*tag= o
FT                   /product= "Polyketide synthase multienzyme housing
FT                   loading module encoded by S. nodosus amphA gene"
FT   CDS             70366. .79938
FT                   /*tag= p
FT                   /product= "Polyketide synthase multienzyme housing
FT                   extension modules 1 and 2 encoded by S. nodosus by amphB
FT                   gene"
FT   CDS             79956. .112709
FT                   /*tag= q
FT                   /product= "Polyketide synthase multienzyme housing
FT                   extension modules 3, 4, 5, 6, 7 and 8 encoded by S.
FT                   nodosus by amphC gene"
XX
PN   WO200297082-A2.
XX
PD   05-DEC-2002.
XX
PF   27-MAY-2002; 2002WO-IE000071.
XX
PR   31-MAY-2001; 2001IE-00000527.
XX
PA   (UYDU-) UNIV COLLEGE DUBLIN.
XX
PI   Caffrey JP;
XX
DR   WPI; 2003-201271/19.
DR   P-PSDB; AAE36116, AAE36117, AAE36118, AAE36119, AAE36120, AAE36121,
DR   AAE36122, AAE36123, AAE36124, AAE36125, AAE36126, AAE36127, AAE36128,
DR   AAE36129, AAE36130, AAE36131, AAE36132.
XX
PT   Novel cytochrome P450 enzyme and nucleotides encoding the enzyme, useful
PT   for preparing amphotericin derivative or analog antibiotic agent with
PT   altered properties, in biosynthesis of polyketide other than
PT   amphotericin.
XX
PS   Claim 1; Page 52-114; 276pp; English.
XX
CC   The invention relates to the gene cluster encoding the polypeptides
```

```
CC   responsible for the biosynthesis of the polyene antibiotic amphotericin
CC   (amph) of Streptomyces nodosus. Polynucleotides of the invention are
CC   useful for preparing amphotericin derivatives or analogue antibiotic
CC   agents with altered properties and in the biosynthesis of polyketides
CC   other than amphotericin. amphDIII, amphDII or amphDI mutants are useful
CC   for producing amphotericin derivatives glycosylated with alternative
CC   sugars; amphDIII or amphDII gene sequences are useful in engineered
CC   biosynthesis of perosaminyl-amphoteronolide B; amphDIII or amphDII and
CC   amphN gene sequences are useful in the engineered biosynthesis of
CC   perosaminyl-16-descarboxyl-16-methyl amphoteronolide B; amphDIII, amphDII
CC   and amphDI gene sequences are useful for preparing polypeptides capable
CC   of addition of mycosamine to a polyketide other than amphoteronolide A or
CC   B or for preparing polypeptides for in vitro synthesis of GDP-mycosamine.
CC   The present sequence is S. nodosus amph biosynthetic gene cluster
XX
SQ   Sequence 113193 BP; 14248 A; 45141 C; 38354 G; 15450 T; 0 U; 0 Other;

  Query Match             5.1%;  Score 52.4;  DB 7;  Length 113193;
  Best Local Similarity   46.9%;  Pred. No. 0.056;
  Matches  164;  Conservative    0;  Mismatches  186;  Indels    0;  Gaps    0;

Qy        320 TCCTCCAGGAGATTGCCCGCTCCACCATCCCCTACGGCAACCGGGAGCTCTGGAGGAAGG 379
              ||  || | |  | ||| ||     |  | | | ||||| |  |  | ||| |      ||
Db      15513 TCGCCCTGCACCTGGCCTGCCAGGCGCTGCGCAACGGCGAGAGCAACCTCGCGGTCGCGG 15572

Qy        380 TGGGGACGGTCGTCTTCATGGTCCCCCTGGAGATGTTGGCCCTCAACCTGGGGGTCACCC 439
                ||   |  |||| | |    |||   | | |  |||| ||| ||  |  | | |||
Db      15573 GCGGCGCCACCGTCATGACCACCCCCAACGCGTTCGTGGCGTTCAGCCGGCAGCGCGCCC 15632

Qy        440 GGCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGAAAAAGGGCCTGGTGGCCACCGACG 499
                | | |  || ||| | ||   ||||     | |    || ||  || ||
Db      15633 TCGCGCAGGACGGCCGGTGCAAGGCGTTCTCCGAGAGCGCCGACGGCATGACCCTCGCCG 15692

Qy        500 TCCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCATCGGCACCCTTTGGGCCGTCCGGC 559
                |  |   | ||||  ||  ||||| | | ||  || | ||    | |  || | |
Db      15693 AAGGTGTCGGCATCGTCCTCGTCGAGCGGCTGTCCGACGCCCGCCGCAACGGCCACCCGG 15752

Qy        560 TGAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACTACATCTACCCCTGGAGGAACCTCG 619
              ||    | | ||    | |  | || |  || |  ||  |  | ||    |  |  ||
Db      15753 TGCTCGCCGTGATCCGCGGCTCCGCCATCAACCAGGACGGCGCCTCCAACGGCCTCAGCG 15812

Qy        620 CCCTAGACATGGCCAACGGCGTGCTCTCCTTCAACTGGGTCAAGGCCTAC 669
              |||   ||    ||  |    || | | ||  | || | |||| ||
Db      15813 CCCCCAACGGCCCCTCCCAGCAGCGCGTCATCCGCCAGGCCCTGGCCAAC 15862


RESULT 15
AAD09400
ID   AAD09400 standard; cDNA; 1815 BP.
XX
AC   AAD09400;
XX
DT   10-SEP-2001  (first entry)
XX
DE   Zea mays neoxanthin cleavage enzyme, VP14 cDNA.
XX
KW   Maize; neoxanthin cleavage enzyme; VP14; abscisic acid; ABA; herbicide;
KW   stress tolerance; transgenic plant; plant breeding; antisense-therapy;
KW   plant growth protectant; ss.
XX
OS   Zea mays.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .1815
FT                   /*tag=  a
FT                   /product= "Zea mays VP14 protein"
XX
PN   EP1116794-A2.
XX
PD   18-JUL-2001.
XX
PF   11-JAN-2001; 2001EP-00300218.
XX
PR   13-JAN-2000; 2000JP-00010056.
PR   11-JAN-2001; 2001JP-00003476.
XX
PA   (RIKE ) RIKEN KK.
XX
PI   Iuchi S,  Kobayashi M,  Shinozaki K;
XX
DR   WPI; 2001-400081/43.
DR   P-PSDB; AAE04788.
XX
PT   A DNA encoding a protein with a neoxanthin cleavage activity for
PT   producing transgenic plants with improved or decreased stress tolerance.
XX
PS   Claim 3; Page 60-64; 101pp; English.
XX
CC   The invention relates to neoxanthin cleavage enzymes and their
CC   corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key role
CC   in endogenous abscisic acid (ABA) biosynthesis under drought stress.
CC   Neoxanthin cleavage enzyme is used for improving stress tolerance in a
CC   plant when expressed in a plant cell. The invention also relates to
CC   methods for increasing or decreasing stress tolerance in a plant by
CC   introducing the DNA into the plant, and a transgenic plant into which a
CC   neoxanthin cleavage enzyme is introduced. The improvement of stress
CC   tolerance in plants is useful, for example in plant breeding. Neoxanthin
CC   cleavage enzyme genes are useful for producing transgenic plants. An arid
CC   land can be improved by growing transformant weed for several years and
CC   then removing the weed by specifically lowering stress tolerance in the
CC   weed by inducing an inducible promoter. The present cDNA sequence encodes
CC   Zea mays neoxanthin cleavage enzyme, VP14 protein related to the
CC   invention
XX
SQ   Sequence 1815 BP; 266 A; 689 C; 604 G; 256 T; 0 U; 0 Other;

  Query Match             5.0%;  Score 51.6;  DB 5;  Length 1815;
  Best Local Similarity   45.9%;  Pred. No. 0.031;
  Matches  177;  Conservative    0;  Mismatches  209;  Indels    0;  Gaps    0;

Qy        320 TCCTCCAGGAGATTGCCCGCTCCACCATCCCCTACGGCAACCGGGAGCTCTGGAGGAAGG 379
              ||  | |   |  ||| |  ||||  |  |||  || ||      |  |   ||      |
Db        155 TCGCCGACCTGCCTGCGCCGTCCAGGAAGCCCGCCGCCATTGCCGTCCCAGGGCACGCCG 214

Qy        380 TGGGGACGGTCGTCTTCATGGTCCCCCTGGAGATGTTGGCCCTCAACCTGGGGGTCACCC 439
              || | ||                || |  |  | ||  | |  | ||    || | | | | |
Db        215 CGGCGCCGAGGAAAGCGGAGGGCGGCAAGAAGCAGCTCAACTTGTTCCAGCGCGCCGCGG 274

Qy        440 GGCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGAAAAAGGGCCTGGTGGCCACCGACG 499
              |  |  || |||| |   || | |   | | |   || | ||| | |   || |||
Db        275 CGGCCGCGCTCGACGCGTTCGAGGAAGGGTTCGTGGCCAACGTCCTCGAGCGGCCCCACG 334

Qy        500 TCCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCATCGGCACCCTTTGGGCCGTCCGGC 559
              || | |   || | | ||  ||     | |  |  |||| | |   | ||||| ||
Db        335 GGCTGCCCAGCACGGCCGACCCGGCCGTGCAGATCGCCGGCAACTTCGCGCCCGTCGGGG 394

Qy        560 TGAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACTACATCTACCCCTGGAGGAACCTCG 619
              ||||||          |  |||| ||  |   || |  ||||   ||||  |   ||   |
Db        395 AGAGGCCGCCCGTGCACGAGCTCCCCGTCTCCGGCCGCATCCCGCCCTTCATCGACGGGG 454

Qy        620 CCCTAGACATGGCCAACGGCGTGCTCTCCTTCAACTGGGTCAAGGCCTACCAGGACCACG 679
              |   |        |  ||  |  | || |||| ||   |||  ||   ||||   |  ||
Db        455 TCTACGCGCGCAACGGCGCCAACCCCTGCTTCGACCCCGTCGCGGGGCACCACCTCTTCG 514

Qy        680 GAATCCGCCCCACCCTGGACGTGCTG 705
                |  |  ||  || ||| ||||
Db        515 ACGGCGACGGCATGGTGCACGCGCTG 540


Search completed: March 17, 2004, 21:47:59
Job time : 255.717 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: March 17, 2004, 21:15:22 ; Search time 47.9049 Seconds
(without alignments)
11885.630 Million cell updates/sec

Title: US-09-664-186-4
Perfect score: 1026
Sequence: 1 gtgaagaacgaaaaaacctt.....ccctcaaggaggcctcctga 1026

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1026	100.0	1026	3	US-09-134-246-4	Sequence 4, Appli
c	2	1026	100.0	5849	3	US-09-134-246-6	Sequence 6, Appli
	3	51.8	5.0	1551	4	US-09-252-991A-5309	Sequence 5309, Ap
	4	51.8	5.0	1563	4	US-09-252-991A-15934	Sequence 15934, A
c	5	51.8	5.0	1953	4	US-09-252-991A-15763	Sequence 15763, A
	6	51.8	5.0	3351	4	US-09-252-991A-15871	Sequence 15871, A
	7	51.8	5.0	4158	4	US-09-252-991A-5348	Sequence 5348, Ap
c	8	51.8	5.0	4953	4	US-09-252-991A-5227	Sequence 5227, Ap
c	9	49.6	4.8	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c	10	49.6	4.8	4411529	3	US-09-103-840A-1	Sequence 1, Appli
	11	46.8	4.6	2460	4	US-09-252-991A-2998	Sequence 2998, Ap
c	12	46.8	4.6	2856	4	US-09-252-991A-2869	Sequence 2869, Ap
	13	46.8	4.6	3387	4	US-09-252-991A-3101	Sequence 3101, Ap
c	14	46.2	4.5	1473	4	US-09-152-060-43	Sequence 43, Appl
	15	46	4.5	1572	4	US-09-252-991A-8725	Sequence 8725, Ap
	16	46	4.5	1929	4	US-09-252-991A-8827	Sequence 8827, Ap
c	17	46	4.5	2079	4	US-09-252-991A-9222	Sequence 9222, Ap
	18	45.8	4.5	960	4	US-09-252-991A-6990	Sequence 6990, Ap
c	19	45.8	4.5	1275	4	US-09-252-991A-7059	Sequence 7059, Ap
	20	45.8	4.5	1926	4	US-09-249-585A-4	Sequence 4, Appli
	21	45.8	4.5	1931	2	US-09-130-114-2	Sequence 2, Appli
	22	45.8	4.5	2187	4	US-09-252-991A-7042	Sequence 7042, Ap
	23	45.6	4.4	984	4	US-09-252-991A-10162	Sequence 10162, A
	24	45.6	4.4	1491	4	US-09-252-991A-10097	Sequence 10097, A
	25	45.6	4.4	2668	2	US-08-461-775-11	Sequence 11, Appl
	26	45.6	4.4	2668	3	US-09-031-606-11	Sequence 11, Appl
c	27	45.4	4.4	1758	4	US-09-252-991A-490	Sequence 490, App
	28	45.4	4.4	1794	4	US-09-252-991A-568	Sequence 568, App
c	29	45.4	4.4	3150	4	US-09-252-991A-11487	Sequence 11487, A
	30	45.4	4.4	3372	4	US-09-252-991A-11405	Sequence 11405, A
c	31	45.2	4.4	1215	4	US-09-252-991A-6208	Sequence 6208, Ap
	32	45.2	4.4	1338	4	US-09-252-991A-6341	Sequence 6341, Ap
	33	45.2	4.4	2034	4	US-09-252-991A-11958	Sequence 11958, A
	34	45.2	4.4	2493	4	US-09-252-991A-11987	Sequence 11987, A
c	35	45.2	4.4	2991	4	US-09-252-991A-12025	Sequence 12025, A
	36	45	4.4	4767	4	US-09-231-899-76	Sequence 76, Appl
	37	44.8	4.4	837	4	US-09-252-991A-4221	Sequence 4221, Ap
c	38	44.8	4.4	1080	4	US-09-252-991A-4664	Sequence 4664, Ap
	39	44.8	4.4	1440	4	US-09-252-991A-4726	Sequence 4726, Ap
	40	44.8	4.4	2458	4	US-09-513-783A-175	Sequence 175, App
	41	44.8	4.4	3060	4	US-09-252-991A-4686	Sequence 4686, Ap
c	42	44.8	4.4	3456	4	US-09-252-991A-4757	Sequence 4757, Ap
c	43	44.8	4.4	23673	4	US-09-773-816-1	Sequence 1, Appli
	44	44	4.3	960	4	US-09-252-991A-11385	Sequence 11385, A
	45	44	4.3	4451	3	US-08-717-294-42	Sequence 42, Appl

ALIGNMENTS

RESULT 1
US-09-134-246-4
; Sequence 4, Application US/09134246B
; Patent No. 6207377
; GENERAL INFORMATION:
; APPLICANT: Wayne, Jay
; APPLICANT: Xu, Shuang-yong
; TITLE OF INVENTION: Method For Construction Of Thermus-E. coli Shuttle
; TITLE OF INVENTION: Vectors And Identification of Two Thermus Plasmid
; TITLE OF INVENTION: Replication Origins
; FILE REFERENCE: Thermus Shuttle Vector
; CURRENT APPLICATION NUMBER: US/09/134,246B
; CURRENT FILING DATE: 1998-08-14
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1026
; TYPE: DNA
; ORGANISM: Thermus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1023)
US-09-134-246-4

Query Match 100.0%; Score 1026; DB 3; Length 1026;
Best Local Similarity 100.0%; Pred. No. 8.3e-241;
Matches 1026; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

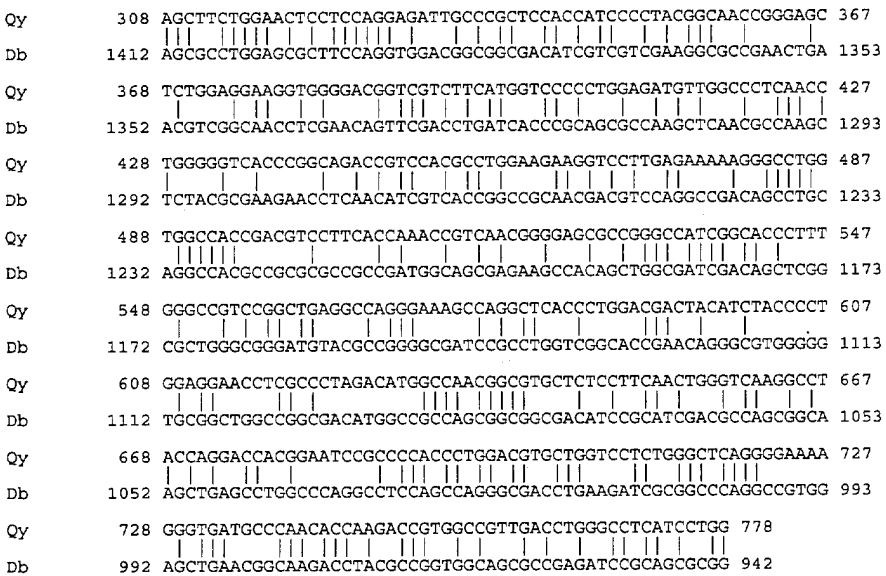
```
Qy          1 GTGAAGAACGAAAAAACCTTCTTTGAAGAGCTTTACGAGGCTTTAGAGGAAACCCACGAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTGAAGAACGAAAAAACCTTCTTTGAAGAGCTTTACGAGGCTTTAGAGGAAACCCACGAC 60

Qy         61 AACACCGATGCCACTAGGGGGTCAGATAGGGGGTCAGAGGACTTCTTCTTGGCCACCGAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACACCGATGCCACTAGGGGGTCAGATAGGGGGTCAGAGGACTTCTTCTTGGCCACCGAC 120

Qy        121 CCCCCTCCAGATGGAGGTGCCGAAAATCGCCTCGCGAAGGGCTTTACATACCAAAAAGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCCCTCCAGATGGAGGTGCCGAAAATCGCCTCGCGAAGGGCTTTACATACCAAAAAGAG 180

Qy        181 GCACTTAGGATTGCTTTACCCGAGAAAGACCATGAGGCTTTCCTTTCCTCTGTTGGGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GCACTTAGGATTGCTTTACCCGAGAAAGACCATGAGGCTTTCCTTTCCTCTGTTGGGGCC 240

Qy        241 CCCCCTATACCACCAGCTGAACCCCCCGTTGGGAATGTATGTCAAGCCGTCCAGGACGGG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCCCCTATACCACCAGCTGAACCCCCCGTTGGGAATGTATGTCAAGCCGTCCAGGACGGG 300

Qy        301 CCTCAGAAGCTTCTGGAACTCCTCCAGGAGATTGCCCGCTCCACCATCCCCTACGGCAAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        301 CCTCAGAAGCTTCTGGAACTCCTCCAGGAGATTGCCCGCTCCACCATCCCCTACGGCAAC 360

Qy        361 CGGGAGCTCTGGAGGAAGGTGGGGACGGTCGTCTTCATGGTCCCCCTGGAGATGTTGGCC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CGGGAGCTCTGGAGGAAGGTGGGGACGGTCGTCTTCATGGTCCCCCTGGAGATGTTGGCC 420

Qy        421 CTCAACCTGGGGGTCACCCGGCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGAAAAAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTCAACCTGGGGGTCACCCGGCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGAAAAAG 480

Qy        481 GGCCTGGTGGCCACCGACGTCCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCATCGGC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GGCCTGGTGGCCACCGACGTCCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCATCGGC 540

Qy        541 ACCCTTTGGGCCGTCCGGCTGAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACTACATC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ACCCTTTGGGCCGTCCGGCTGAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACTACATC 600

Qy        601 TACCCCTGGAGGAACCTCGCCCTAGACATGGCCAACGGCGTGCTCTCCTTCAACTGGGTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TACCCCTGGAGGAACCTCGCCCTAGACATGGCCAACGGCGTGCTCTCCTTCAACTGGGTC 660

Qy        661 AAGGCCTACCAGGACCACGGAATCCGCCCCACCCTGGACGTGCTGGTCCTCTGGGCTCAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AAGGCCTACCAGGACCACGGAATCCGCCCCACCCTGGACGTGCTGGTCCTCTGGGCTCAG 720

Qy        721 GGGAAAAGGGTGATGCCCAACACCAAGACCGTGGCCGTTGACCTGGGCCTCATCCTGGTC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GGGAAAAGGGTGATGCCCAACACCAAGACCGTGGCCGTTGACCTGGGCCTCATCCTGGTC 780

Qy        781 CTCCCCGAGGTGGAGCGTTCCAAACTCCCGGCCCTTATCACCCTCATTGCTACGTACATT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CTCCCCGAGGTGGAGCGTTCCAAACTCCCGGCCCTTATCACCCTCATTGCTACGTACATT 840

Qy        841 GCCGATCTCCTAGATGACCGTCGTTCAAGACGTTTCTATGCAGGCTTGCTGTGGGCTGTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCCGATCTCCTAGATGACCGTCGTTCAAGACGTTTCTATGCAGGCTTGCTGTGGGCTGTG 900

Qy        901 GCCAGGGGTGAACTCCCCGCGCAATATCTATTTGCCGTCCTAATGCGGGTTATCCGAGAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GCCAGGGGTGAACTCCCCGCGCAATATCTATTTGCCGTCCTAATGCGGGTTATCCGAGAT 960

Qy        961 TACACGGATGGCCATCTGACACGACCGGGAGCGTACCTAGTGAAGACCCTCAAGGAGGCC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TACACGGATGGCCATCTGACACGACCGGGAGCGTACCTAGTGAAGACCCTCAAGGAGGCC 1020

Qy       1021 TCCTGA 1026
              ||||||
Db       1021 TCCTGA 1026
```

```
RESULT 2
US-09-134-246-6/c
; Sequence 6, Application US/09134246B
; Patent No. 6207377
; GENERAL INFORMATION:
;  APPLICANT: Wayne, Jay
;  APPLICANT: Xu, Shuang-yong
;  TITLE OF INVENTION: Method For Construction Of Thermus-E. coli Shuttle
;  TITLE OF INVENTION: Vectors And Identification of Two Thermus Plasmid
;  TITLE OF INVENTION: Replication Origins
;  FILE REFERENCE: Thermus Shuttle Vector
;  CURRENT APPLICATION NUMBER: US/09/134,246B
;  CURRENT FILING DATE: 1998-08-14
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
;   LENGTH: 5849
;   TYPE: DNA
;   ORGANISM: Thermus sp.
US-09-134-246-6

  Query Match           100.0%;  Score 1026;  DB 3;  Length 5849;
```

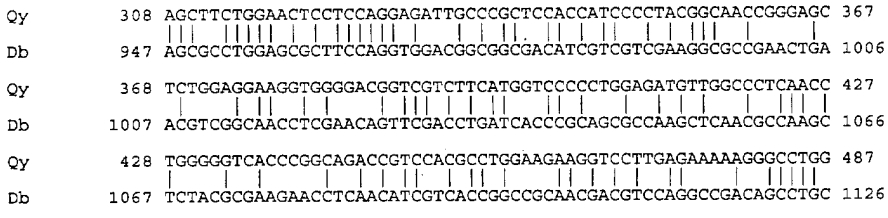
```
 Best Local Similarity   100.0%;  Pred. No. 1.3e-240;
 Matches 1026;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTGAAGAACGAAAAAACCTTCTTTGAAGAGCTTTACGAGGCTTTAGAGGAAACCCACGAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4157 GTGAAGAACGAAAAAACCTTCTTTGAAGAGCTTTACGAGGCTTTAGAGGAAACCCACGAC 4098

Qy         61 AACACCGATGCCACTAGGGGGTCAGATAGGGGGTCAGAGGACTTCTTCTTGGCCACCGAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4097 AACACCGATGCCACTAGGGGGTCAGATAGGGGGTCAGAGGACTTCTTCTTGGCCACCGAC 4038

Qy        121 CCCCCTCCAGATGGAGGTGCCGAAAATCGCCTCGCGAAGGGCTTTACATACCAAAAAGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4037 CCCCCTCCAGATGGAGGTGCCGAAAATCGCCTCGCGAAGGGCTTTACATACCAAAAAGAG 3978

Qy        181 GCACTTAGGATTGCTTTACCCGAGAAAGACCATGAGGCTTTCCTTTCCTCTGTTGGGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3977 GCACTTAGGATTGCTTTACCCGAGAAAGACCATGAGGCTTTCCTTTCCTCTGTTGGGGCC 3918

Qy        241 CCCCCTATACCACCAGCTGAACCCCCCGTTGGGAATGTATGTCAAGCCGTCCAGGACGGG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3917 CCCCCTATACCACCAGCTGAACCCCCCGTTGGGAATGTATGTCAAGCCGTCCAGGACGGG 3858

Qy        301 CCTCAGAAGCTTCTGGAACTCCTCCAGGAGATTGCCCGCTCCACCATCCCCTACGGCAAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3857 CCTCAGAAGCTTCTGGAACTCCTCCAGGAGATTGCCCGCTCCACCATCCCCTACGGCAAC 3798

Qy        361 CGGGAGCTCTGGAGGAAGGTGGGGACGGTCGTCTTCATGGTCCCCCTGGAGATGTTGGCC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3797 CGGGAGCTCTGGAGGAAGGTGGGGACGGTCGTCTTCATGGTCCCCCTGGAGATGTTGGCC 3738

Qy        421 CTCAACCTGGGGGTCACCCGGCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGAAAAAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3737 CTCAACCTGGGGGTCACCCGGCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGAAAAAG 3678

Qy        481 GGCCTGGTGGCCACCGACGTCCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCATCGGC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3677 GGCCTGGTGGCCACCGACGTCCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCATCGGC 3618

Qy        541 ACCCTTTGGGCCGTCCGGCTGAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACTACATC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3617 ACCCTTTGGGCCGTCCGGCTGAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACTACATC 3558

Qy        601 TACCCCTGGAGGAACCTCGCCCTAGACATGGCCAACGGCGTGCTCTCCTTCAACTGGGTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3557 TACCCCTGGAGGAACCTCGCCCTAGACATGGCCAACGGCGTGCTCTCCTTCAACTGGGTC 3498

Qy        661 AAGGCCTACCAGGACCACGGAATCCGCCCCACCCTGGACGTGCTGGTCCTCTGGGCTCAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3497 AAGGCCTACCAGGACCACGGAATCCGCCCCACCCTGGACGTGCTGGTCCTCTGGGCTCAG 3438

Qy        721 GGGAAAAGGGTGATGCCCAACACCAAGACCGTGGCCGTTGACCTGGGCCTCATCCTGGTC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3437 GGGAAAAGGGTGATGCCCAACACCAAGACCGTGGCCGTTGACCTGGGCCTCATCCTGGTC 3378

Qy        781 CTCCCCGAGGTGGAGCGTTCCAAACTCCCGGCCCTTATCACCCTCATTGCTACGTACATT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3377 CTCCCCGAGGTGGAGCGTTCCAAACTCCCGGCCCTTATCACCCTCATTGCTACGTACATT 3318

Qy        841 GCCGATCTCCTAGATGACCGTCGTTCAAGACGTTTCTATGCAGGCTTGCTGTGGGCTGTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3317 GCCGATCTCCTAGATGACCGTCGTTCAAGACGTTTCTATGCAGGCTTGCTGTGGGCTGTG 3258

Qy        901 GCCAGGGGTGAACTCCCCGCGCAATATCTATTTGCCGTCCTAATGCGGGTTATCCGAGAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3257 GCCAGGGGTGAACTCCCCGCGCAATATCTATTTGCCGTCCTAATGCGGGTTATCCGAGAT 3198

Qy        961 TACACGGATGGCCATCTGACACGACCGGGAGCGTACCTAGTGAAGACCCTCAAGGAGGCC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3197 TACACGGATGGCCATCTGACACGACCGGGAGCGTACCTAGTGAAGACCCTCAAGGAGGCC 3138

Qy       1021 TCCTGA 1026
              ||||||
```

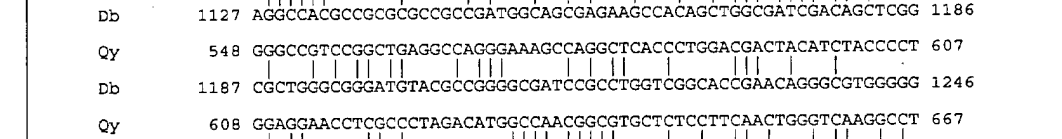
```
Db       3137 TCCTGA 3132


RESULT 3
US-09-252-991A-5309
; Sequence 5309, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 5309
;   LENGTH: 1551
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5309

  Query Match             5.0%;  Score 51.8;  DB 4;  Length 1551;
  Best Local Similarity   44.4%;  Pred. No. 0.0017;
  Matches  209;  Conservative    0;  Mismatches  262;  Indels    0;  Gaps    0;

Qy        308 AGCTTCTGGAACTCCTCCAGGAGATTGCCCGCTCCACCATCCCCTACGGCAACCGGGAGC 367
              |||  |||||  |  |  ||||||  |   |  |  ||  ||  |  |  |||  |
Db        507 AGCGCCTGGAGCGCTTCCAGGTGGACGGCGGCGACATCGTCGTCGAAGGCGCCGAACTGA 566

Qy        368 TCTGGAGGAAGGTGGGGACGGTCGTCTTCATGGTCCCCCTGGAGATGTTGGCCCTCAACC 427
              |    |  ||  |  |    | ||| | | ||   ||  |   |  |  |   |  ||| |
Db        567 ACGTCGGCAACCTCGAACAGTTCGACCTGATCACCCGCAGCGCCAAGCTCAACGCCAAGC 626

Qy        428 TGGGGGTCACCCGGCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGAAAAAGGGCCTGG 487
              |     |  |      |  | | ||  | ||  |    ||  | |  |  ||    |  |||||
Db        627 TCTACGCGAAGAACCTCAACATCGTCACCGGCCGCAACGACGTCCAGGCCGACAGCCTGC 686

Qy        488 TGGCCACCGACGTCCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCATCGGCACCCTTT 547
              ||||||       |    | |   | | |    |   | |  ||| |||| || |
Db        687 AGGCCACGCCGCGCGCCGCCGATGGCAGCGAGAAGCCACAGCTGGCGATCGACAGCTCGG 746

Qy        548 GGGCCGTCCGGCTGAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACTACATCTACCCCT 607
              |   | | || ||    | |||       | ||  |   |||  |  |
Db        747 CGCTGGGCGGGATGTACGCCGGGGCGATCCGCCTGGTCGGCACCGAGCAGGGCGTGGGGG 806

Qy        608 GGAGGAACCTCGCCCTAGACATGGCCAACGGCGTGCTCTCCTTCAACTGGGTCAAGGCCT 667
              | ||     ||  |          |||| ||||     |  | | |   | ||  | ||
Db        807 TGCGGCTGGCCGGCGACATGGCCGCCAGCGGCGGCGACATCCGCATCGACGCCAGCGGCA 866

Qy        668 ACCAGGACCACGGAATCCGCCCCACCCTGGACGTGCTGGTCCTCTGGGCTCAGGGGAAAA 727
              | | |  ||  |       | ||| || || ||  |||     ||  ||| ||||
Db        867 AGCTGAGCCTGGCCCAGGCCTCCAGCCAGGGCGACCTGAAGATCGCGGCCCAGGCCGTGG 926

Qy        728 GGGTGATGCCCAACACCAAGACCGTGGCCGTTGACCTGGGCCTCATCCTGG 778
              | |||    ||| ||| |  |||  | |   |  | || ||| |  ||
Db        927 AGCTGAACGGCAAGACCTACGCCGGTGGCAGCGCCGAGATCCGCAGCGCGG 977


RESULT 4
US-09-252-991A-15934
; Sequence 15934, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 15934
;   LENGTH: 1563
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-15934

  Query Match             5.0%;  Score 51.8;  DB 4;  Length 1563;
  Best Local Similarity   44.4%;  Pred. No. 0.0017;
  Matches  209;  Conservative    0;  Mismatches  262;  Indels    0;  Gaps    0;

Qy        308 AGCTTCTGGAACTCCTCCAGGAGATTGCCCGCTCCACCATCCCCTACGGCAACCGGGAGC 367
              |||  |||||  |  |  ||||||  |   |  |  ||  ||  |  |  |||  |
Db        507 AGCGCCTGGAGCGCTTCCAGGTGGACGGCGGCGACATCGTCGTCGAAGGCGCCGAACTGA 566

Qy        368 TCTGGAGGAAGGTGGGGACGGTCGTCTTCATGGTCCCCCTGGAGATGTTGGCCCTCAACC 427
              |    |  ||  |  |    | ||| | | ||   ||  |   |  |  |   |  ||| |
Db        567 ACGTCGGCAACCTCGAACAGTTCGACCTGATCACCCGCAGCGCCAAGCTCAACGCCAAGC 626

Qy        428 TGGGGGTCACCCGGCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGAAAAAGGGCCTGG 487
              |     |  |      |  | | ||  | ||  |    ||  | |  |  ||    |  |||||
Db        627 TCTACGCGAAGAACCTCAACATCGTCACCGGCCGCAACGACGTCCAGGCCGACAGCCTGC 686

Qy        488 TGGCCACCGACGTCCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCATCGGCACCCTTT 547
              ||||||       |    | |   | | |    |   | |  ||| |||| || |
Db        687 AGGCCACGCCGCGCGCCGCCGATGGCAGCGAGAAGCCACAGCTGGCGATCGACAGCTCGG 746

Qy        548 GGGCCGTCCGGCTGAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACTACATCTACCCCT 607
              |   | | || ||    | |||       | ||  |   |||  |  |
Db        747 CGCTGGGCGGGATGTACGCCGGGGCGATCCGCCTGGTCGGCACCGAGCAGGGCGTGGGGG 806

Qy        608 GGAGGAACCTCGCCCTAGACATGGCCAACGGCGTGCTCTCCTTCAACTGGGTCAAGGCCT 667
              | ||     ||  |          |||| ||||     |  | | |   | ||  | ||
Db        807 TGCGGCTGGCCGGCGACATGGCCGCCAGCGGCGGCGACATCCGCATCGACGCCAGCGGCA 866

Qy        668 ACCAGGACCACGGAATCCGCCCCACCCTGGACGTGCTGGTCCTCTGGGCTCAGGGGAAAA 727
              | | |  ||  |       | ||| || || ||  |||     ||  ||| ||||
Db        867 AGCTGAGCCTGGCCCAGGCCTCCAGCCAGGGCGACCTGAAGATCGCGGCCCAGGCCGTGG 926

Qy        728 GGGTGATGCCCAACACCAAGACCGTGGCCGTTGACCTGGGCCTCATCCTGG 778
              | |||    ||| ||| |  |||  | |   |  | || ||| |  ||
Db        927 AGCTGAACGGCAAGACCTACGCCGGTGGCAGCGCCGAGATCCGCAGCGCGG 977


RESULT 5
US-09-252-991A-15763/c
; Sequence 15763, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 15763
;   LENGTH: 1953
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-15763

  Query Match             5.0%;  Score 51.8;  DB 4;  Length 1953;
  Best Local Similarity   44.4%;  Pred. No. 0.0018;
  Matches  209;  Conservative    0;  Mismatches  262;  Indels    0;  Gaps    0;
```



RESULT 6
US-09-252-991A-15871
; Sequence 15871, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 15871
; LENGTH: 3351
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-15871

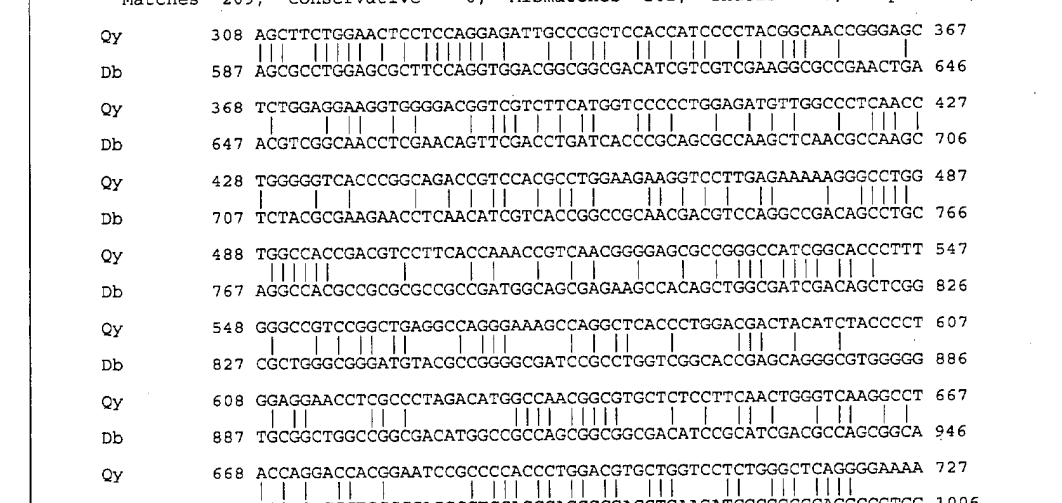
Query Match 5.0%; Score 51.8; DB 4; Length 3351;
Best Local Similarity 44.4%; Pred. No. 0.0021;
Matches 209; Conservative 0; Mismatches 262; Indels 0; Gaps 0;





RESULT 7
US-09-252-991A-5348
; Sequence 5348, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 5348
; LENGTH: 4158
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5348

Query Match 5.0%; Score 51.8; DB 4; Length 4158;
Best Local Similarity 44.4%; Pred. No. 0.0022;
Matches 209; Conservative 0; Mismatches 262; Indels 0; Gaps 0;



```
Qy        728 GGGTGATGCCCAACACCAAGACCGTGGCCGTTGACCTGGGCCTCATCCTGG 778
Db       1007 AGCTGAACGGCAAGACCTACGCCGGTGGCAGCGCCGAGATCCGCAGCGCGG 1057

RESULT 8
US-09-252-991A-5227/c
; Sequence 5227, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 5227
;   LENGTH: 4953
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5227

  Query Match             5.0%;  Score 51.8;  DB 4;  Length 4953;
  Best Local Similarity   44.4%;  Pred. No. 0.0023;
  Matches  209;  Conservative    0;  Mismatches  262;  Indels    0;  Gaps    0;

Qy        308 AGCTTCTGGAACTCCTCCAGGAGATTGCCCGCTCCACCATCCCCTACGGCAACCGGGAGC 367
Db       4412 AGCGCCTGGAGCGCTTCCAGGTGGACGGCGGCGACATCGTCGTCGAAGGCGCCGAACTGA 4353

Qy        368 TCTGGAGGAAGGTGGGGACGGTCGTCTTCATGGTCCCCCTGGAGATGTTGGCCCTCAACC 427
Db       4352 ACGTCGGCAACCTCGAACAGTTCGACCTGATCACCCGCAGCGCCAAGCTCAACGCCAAGC 4293

Qy        428 TGGGGGTCACCCGGCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGAAAAAGGGCCTGG 487
Db       4292 TCTACGCGAAGAACCTCAACATCGTCACCGGCCGCAACGACGTCCAGGCCGACAGCCTGC 4233

Qy        488 TGGCCACCGACGTCCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCATCGGCACCCTTT 547
Db       4232 AGGCCACGCCGCGCGCCGCCGATGGCAGCGAGAAGCCACAGCTGGCGATCGACAGCTCGG 4173

Qy        548 GGGCCGTCCGGCTGAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACTACATCTACCCCT 607
Db       4172 CGCTGGGCGGGATGTACGCCGGGGCGATCCGCCTGGTCGGCACCGAGCAGGGCGTGGGGG 4113

Qy        608 GGAGGAACCTCGCCCTAGACATGGCCAACGGCGTGCTCTCCTTCAACTGGGTCAAGGCCT 667
Db       4112 TGCGGCTGGCCGGCGACATGGCCGCCAGCGGCGGCGACATCCGCATCGACGCCAGCGGCA 4053

Qy        668 ACCAGGACCACGGAATCCGCCCCACCCTGGACGTGCTGGTCCTCTGGGCTCAGGGGAAAA 727
Db       4052 AGCTGAGCCTGGCCCAGGCCTCCAGCCAGGGCGACCTGAAGATCGCGGCCCAGGCCGTGG 3993

Qy        728 GGGTGATGCCCAACACCAAGACCGTGGCCGTTGACCTGGGCCTCATCCTGG 778
Db       3992 AGCTGAACGGCAAGACCTACGCCGGTGGCAGCGCCGAGATCCGCAGCGCGG 3942

RESULT 9
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

  Query Match             4.8%;  Score 49.6;  DB 3;  Length 4403765;
  Best Local Similarity   45.7%;  Pred. No. 0.053;
  Matches  212;  Conservative    0;  Mismatches  249;  Indels    3;  Gaps    1;

Qy        391 GTCTTCATGGTCCCCCTGGAGATGTTGGCCCTCAACCTGGGGGTCACCCGGCAGACCGTC 450
Db    2780667 GCCTTCCTGGCCACCGGCGCCACCGTGGGCATCCTGCAACAGGAACCGCCGCTGAACGAG 2780608

Qy        451 CACGCCTGGAAGAAGGTCCTTGAGAAAAAGGGCCTGGTGGCCACCGACGTCCTTCACCAA 510
Db    2780607 GACAAGACCGTTCGCGGCAATGTGGAAGAGGGCATGGGGGACATCAAGATCAAGCTCGAC 2780548

Qy        511 ACCGTCAACGGGGAGCGCCGGGCCATCGGCACCCTTTGGGCCGTCCGGCTGAGGCCAGGG 570
Db    2780547 CGCTTCAACGAGGTCGCCGAATTGATGGCCACCGACTACACCGACGAGCTGATGGAAGAG 2780488

Qy        571 AAAGCCAGGCTCACCCTGGACGACTACATCTACCCCTGGAGGAACCTCGCCCTAGACATG 630
Db    2780487 ATGGGTCGGCTGCAAGAGGAACTGGACCACGCCGACGCGTGGGACCTCGACGCGCAGCTC 2780428

Qy        631 GCCAACGGCGTGCTCTCCTTCAACTGGGTCAAGGCCTACCAGGACCACGGAATCCGCCCC 690
Db    2780427 GAGCAGGCCATGGATGCGCTGCGCTGTCCGCCGGCCGAC---GAGCCGGTAACCAACCTA 2780371

Qy        691 ACCCTGGACGTGCTGGTCCTCTGGGCTCAGGGGAAAAGGGTGATGCCCAACACCAAGACC 750
Db    2780370 TCCGGTGGCGAGCGTCGCCGGGTGGCGCTGTGCAAACTGCTGTTGTCCAAACCCGACCTG 2780311

Qy        751 GTGGCCGTTGACCTGGGCCTCATCCTGGTCCTCCCCGAGGTGGAGCGTTCCAAACTCCCG 810
Db    2780310 TTGTTGCTCGACGAGCCGACCAACCACCTAGACGCGGAAAGTGTGCAGTGGCTCGAACAG 2780251

Qy        811 GCCCTTATCACCCTCATTGCTACGTACATTGCCGATCTCCTAGA 854
Db    2780250 CATCTGGCCAGCTACCCCGGTGCGATCCTGGCGGTCACCCACGA 2780207

RESULT 10
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
```

```
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1

  Query Match             4.8%;  Score 49.6;  DB 3;  Length 4411529;
  Best Local Similarity   45.7%;  Pred. No. 0.053;
  Matches  212;  Conservative    0;  Mismatches  249;  Indels    3;  Gaps    1;

Qy        391 GTCTTCATGGTCCCCCTGGAGATGTTGGCCCTCAACCTGGGGGTCACCCGGCAGACCGTC 450
Db    2783857 GCCTTCCTGGCCACCGGCGCCACCGTGGGCATCCTGCAACAGGAACCGCCGCTGAACGAG 2783798
Qy        451 CACGCCTGGAAGAAGGTCCTTGAGAAAAAGGGCCTGGTGGCCACCGACGTCCTTCACCAA 510
Db    2783797 GACAAGACCGTTCGCGGCAATGTGGAAGAGGGCATGGGGACATCAAGATCAAGCTCGAC 2783738
Qy        511 ACCGTCAACGGGGAGCGCCGGGCCATCGGCACCCTTTGGGCCGTCCGGCTGAGGCCAGGG 570
Db    2783737 CGCTTCAACGAGGTCGCCGAATTGATGGCCACCGACTACACCGACGAGCTGATGGAAGAG 2783678
Qy        571 AAAGCCAGGCTCACCCTGGACGACTACATCTACCCCTGGAGGAACCTCGCCCTAGACATG 630
Db    2783677 ATGGGTCGGCTGCAAGAGGAACTGGACCACGCCGACGCGTGGGACCTCGACGCGCAGCTC 2783618
Qy        631 GCCAACGGCGTGCTCTCCTTCAACTGGGTCAAGGCCTACCAGGACCACGGAATCCGCCCC 690
Db    2783617 GAGCAGGCCATGGATGCGCTGCGCTGTCCGCCGGCCGAC---GAGCCGGTAACCAACCTA 2783561
Qy        691 ACCCTGGACGTGCTGGTCCTCTGGGCTCAGGGGAAAAGGGTGATGCCCAACACCAAGACC 750
Db    2783560 TCCGGTGGCGAGCGTCGCCGGGTGGCGCTGTGCAAACTGCTGTTGTCCAAACCCGACCTG 2783501
Qy        751 GTGGCCGTTGACCTGGGCCTCATCCTGGTCCTCCCCGAGGTGGAGCGTTCCAAACTCCCG 810
Db    2783500 TTGTTGCTCGACGAGCCGACCAACCACCTAGACGCGGAAAGTGTGCAGTGGCTCGAACAG 2783441
Qy        811 GCCCTTATCACCCTCATTGCTACGTACATTGCCGATCTCCTAGA 854
Db    2783440 CATCTGGCCAGCTACCCCGGTGCGATCCTGGCGGTCACCCACGA 2783397

RESULT 11
US-09-252-991A-2998
; Sequence 2998, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J.  Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 2998
;   LENGTH: 2460
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-2998

  Query Match             4.6%;  Score 46.8;  DB 4;  Length 2460;
  Best Local Similarity   46.4%;  Pred. No. 0.032;
  Matches  153;  Conservative    0;  Mismatches  177;  Indels    0;  Gaps    0;

Qy        356 GCAACCGGGAGCTCTGGAGGAAGGTGGGGACGGTCGTCTTCATGGTCCCCCTGGAGATGT 415
Db       1500 GCAACCGCCTGCACTGGGATTTCCGCATGTTCGACCTGTACGGCATCCGCAGCGGTCCCG 1559
Qy        416 TGGCCCTCAACCTGGGGGTCACCCGGCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGA 475
Db       1560 GGGAGACCGCCGTGGAGCAGTGGAAGGCGAGCCTGCATGCCGAGGATCACGACCGCGTGC 1619
Qy        476 AAAAGGGCCTGGTGGCCACCGACGTCCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCA 535
Db       1620 TGCGCGAGCTGGAGCGCGCCGTCAGCGGGTTGCAGAAATTCGACTGCGAGTTCCGTATCG 1679
Qy        536 TCGGCACCCTTTGGGCCGTCCGGCTGAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACT 595
Db       1680 TCCGCCCCAACCGCGAGGTTCGCCACCTGCGGGCGATCGCCACGCTGACGCGCGACGAGG 1739
Qy        596 ACATCTACCCCTGGAGGAACCTCGCCCTAGACATGGCCAACGGCGTGCTCTCCTTCAACT 655
Db       1740 ACAACCGTCCGGTGCGGATGATCGGCATCAACAGCGACATCACCGAGATCCGCACCCTGG 1799
Qy        656 GGGTCAAGGCCTACCAGGACCACGGAATCC 685
Db       1800 CGGAAACCCTGCACGAGGAAAAGGAACGCC 1829

RESULT 12
US-09-252-991A-2869/c
; Sequence 2869, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J.  Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 2869
;   LENGTH: 2856
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-2869

  Query Match             4.6%;  Score 46.8;  DB 4;  Length 2856;
  Best Local Similarity   46.4%;  Pred. No. 0.033;
  Matches  153;  Conservative    0;  Mismatches  177;  Indels    0;  Gaps    0;

Qy        356 GCAACCGGGAGCTCTGGAGGAAGGTGGGGACGGTCGTCTTCATGGTCCCCCTGGAGATGT 415
Db       2027 GCAACCGCCTGCACTGGGATTTCCGCATGTTCGACCTGTACGGCATCCGCAGCGGTCCCG 1968
Qy        416 TGGCCCTCAACCTGGGGGTCACCCGGCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGA 475
Db       1967 GGGAGACCGCCGTGGAGCAGTGGAAGGCGAGCCTGCATGCCGAGGATCACGACCGCGTGC 1908
Qy        476 AAAAGGGCCTGGTGGCCACCGACGTCCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCA 535
Db       1907 TGCGCGAGCTGGAGCGCGCCGTCAGCGGGTTGCAGAAATTCGACTGCGAGTTCCGTATCG 1848
Qy        536 TCGGCACCCTTTGGGCCGTCCGGCTGAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACT 595
Db       1847 TCCGCCCCAACCGCGAGGTTCGCCACCTGCGGGCGATCGCCACGCTGACGCGCGACGAGG 1788
Qy        596 ACATCTACCCCTGGAGGAACCTCGCCCTAGACATGGCCAACGGCGTGCTCTCCTTCAACT 655
Db       1787 ACAACCGTCCGGTGCGGATGATCGGCATCAACAGCGACATCACCGAGATCCGCACCCTGG 1728
Qy        656 GGGTCAAGGCCTACCAGGACCACGGAATCC 685
Db       1727 CGGAAACCCTGCACGAGGAAAAGGAACGCC 1698

RESULT 13
US-09-252-991A-3101
```

; Sequence 3101, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 3101
; LENGTH: 3387
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-3101

Query Match 4.6%; Score 46.8; DB 4; Length 3387;
Best Local Similarity 46.4%; Pred. No. 0.035;
Matches 153; Conservative 0; Mismatches 177; Indels 0; Gaps 0;

Qy 356 GCAACCGGGAGCTCTGGAGGAAGGTGGGGACGGTCGTCTTCATGGTCCCCCTGGAGATGT 415
Db 1364 GCAACCGCCTGCACTGGGATTTCCGCATGTTCGACCTGTACGGCATCCGCAGCGGTCCCG 1423

Qy 416 TGGCCCTCAACCTGGGGGTCACCCGGCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGA 475
Db 1424 GGGAGACCGCCGTGGAGCAGTGGAAGGCGAGCCTGCATGCCGAGGATCACGACCGCGTGC 1483

Qy 476 AAAAGGGCCTGGTGGCCACCGACGTCCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCA 535
Db 1484 TGCGCGAGCTGGAGCGCGCCGTCAGCGGGTTGCAGAAATTCGACTGCGAGTTCCGTATCG 1543

Qy 536 TCGGCACCCTTTGGGCCGTCCGGCTGAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACT 595
Db 1544 TCCGCCCCAACCGCGAGGTTCGCCACCTGCGGGCGATCGCCACGCTGACGCGCGACGAGG 1603

Qy 596 ACATCTACCCCTGGAGGAACCTCGCCCTAGACATGGCCAACGGCGTGCTCTCCTTCAACT 655
Db 1604 ACAACCGTCCGGTGCGGATGATCGGCATCAACAGCGACATCACCGAGATCCGCACCCTGG 1663

Qy 656 GGGTCAAGGCCTACCAGGACCACGGAATCC 685
Db 1664 CGGAAACCCTGCACGAGGAAAAGGAACGCC 1693

RESULT 14
US-09-152-060-43/c
; Sequence 43, Application US/09152060
; Patent No. 6448230
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 28 Human Secreted Proteins
; FILE REFERENCE: PZ003P1.US
; CURRENT APPLICATION NUMBER: US/09/152,060
; CURRENT FILING DATE: 1998-09-11
; EARLIER APPLICATION NUMBER: PCT/US98/04858
; EARLIER FILING DATE: 1998-03-12
; EARLIER APPLICATION NUMBER: 60/040,762
; EARLIER FILING DATE: 1997-03-14
; EARLIER APPLICATION NUMBER: 60/040,710
; EARLIER FILING DATE: 1997-03-14
; EARLIER APPLICATION NUMBER: 60/050,934
; EARLIER FILING DATE: 1997-05-30
; EARLIER APPLICATION NUMBER: 60/048,100
; EARLIER FILING DATE: 1997-05-30
; EARLIER APPLICATION NUMBER: 60/048,357
; EARLIER FILING DATE: 1997-05-30
; EARLIER APPLICATION NUMBER: 60/048,189
; EARLIER FILING DATE: 1997-05-30
; EARLIER APPLICATION NUMBER: 60/057,765
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: 60/048,970
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/068,368
; EARLIER FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 43
; LENGTH: 1473
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-152-060-43

Query Match 4.5%; Score 46.2; DB 4; Length 1473;
Best Local Similarity 42.5%; Pred. No. 0.038;
Matches 299; Conservative 0; Mismatches 403; Indels 1; Gaps 1;

Qy 314 TGGAACTCCTCCAGGAGATTGCCCGCTCCACCATCCCCTACGGCAACCGGGAGCTCTGGA 373
Db 823 TGGAACCTGTCCTGACTATTCCAAATCTCCAGGCCCACCAGGACCACACGGAAGTTGAGT 764

Qy 374 GGAAGGTGGGGACGGTCGTCTTCATGGTCCCCCTGGAGATGTTGGCCCTCAACCTGGGGG 433
Db 763 TTCTGATATAGCTTGTCCACGTGATTCACCACCTCCAGCACCCGATGACGCACGGCTGCT 704

Qy 434 TCACCCGGCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGAAAAAGGGCCTGGTGGCCA 493
Db 703 TCGCTCCCCAGCATCTGGAACTCTGCATTGTCCACGACCACATACAGCTCCACGTAGCGG 644

Qy 494 CCGACGTCCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCATCGGC-ACCCTTTGGGCC 552
Db 643 GTCTCTCGGGATGGCAGAGAGTCCCCGGGCCGAGGCCTGAAGACGGCTGCCGTCCGGGGT 584

Qy 553 GTCCGGCTGAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACTACATCTACCCCTGGAGG 612
Db 583 CCCAGGAGGCTGCCCAGGCTGTCGTCGCTGACCCCGCAGGTCCCGGCCGTCTGCAGCAGG 524

Qy 613 AACCTCGCCCTAGACATGGCCAACGGCGTGCTCTCCTTCAACTGGGTCAAGGCCTACCAG 672
Db 523 TGCTCAGCCTGGTACACGGCGTGCCGTCCGCCCTCGCCACCTTCATCCAGGGGCTCGATC 464

Qy 673 GACCACGGAATCCGCCCCACCCTGGACGTGCTGGTCCTCTGGGCTCAGGGGAAAAGGGTG 732
Db 463 AGGTGCAGGTCTGACCCCACCTGGAAGAAACCCCTGAGGCCGGCACAGGTGCTGAGGCTG 404

Qy 733 ATGCCCAACACCAAGACCGTGGCCGTTGACCTGGGCCTCATCCTGGTCCTCCCCGAGGTG 792
Db 403 GCGGCTGAGTCCGGTACCCTCTAAGTGGCCCTGGTAGAAGCAGTGGTCCTGCCCGCGAGG 344

Qy 793 GAGCGTTCCAAACTCCCGGCCCTTATCACCCTCATTGCTACGTACATTGCCGATCTCCTA 852
Db 343 CTGCTCCGTCACCTCGGAGCCATTGGCAGCCGTATAGGTCTCTGTGTAGCCGGAGCCCAG 284

Qy 853 GATGACCGTCGTTCAAGACGTTTCTATGCAGGCTTGCTGTGGGCTGTGGCCAGGGGTGAA 912
Db 283 CAGGTCCCTGTTCTTCCGCAGGTGGAGGGTGAAGTTGTGCCCTGTGGCCCCAAGGACGTA 224

Qy 913 CTCCCCGCGCAATATCTATTTGCCGTCCTAATGCGGGTTATCCGAGATTACACGGATGGC 972
Db 223 GCTCACCCTCTCTGGGTGCAGGCCCAAGTGGGAGGGCAGAGCTCGGCGGACTCGGGGGCC 164

Qy 973 CATCTGACACGACCGGGAGCGTACCTAGTGAAGACCCTCAAGG 1015
Db 163 TGGCAGACGCCACGGCAACACGACCTCATACTGCTCCATGAGG 121

RESULT 15
US-09-252-991A-8725
; Sequence 8725, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.

```
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 8725
;   LENGTH: 1572
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-8725

  Query Match             4.5%;  Score 46;  DB 4;  Length 1572;
  Best Local Similarity   49.2%;  Pred. No. 0.044;
  Matches  121;  Conservative    0;  Mismatches  125;  Indels    0;  Gaps    0;

Qy        399 GGTCCCCCTGGAGATGTTGGCCCTCAACCTGGGGGTCACCCGGCAGACCGTCCACGCCTG 458
              || |  || |  |  |  ||  | || || ||   ||  |    || ||| |  ||   |
Db        196 GGACGGCCCGCTGGCGCCGGTACGCAGCCCGGTATTCCTCGCCGAGGCCGACGGCGAGCG 255

Qy        459 GAAGAAGGTCCTTGAGAAAAAGGGCCTGGTGGCCACCGACGTCCTTCACCAAACCGTCAA 518
                ||  | |||| |  |  |      ||| | | | |||||| |   | |     |  || |
Db        256 CCAGGTGATCCTCGGCAGCACCCCGCTGCTCGACGCCGACGCCGCGCTCGCCGCGCTCGA 315

Qy        519 CGGGGAGCGCCGGGCCATCGGCACCCTTTGGGCCGTCCGGCTGAGGCCAGGGAAAGCCAG 578
              ||  | |   ||||||  || |  |      || | |  ||| ||  |   ||   |||
Db        316 CGCCGCGGTGCGGGCCTACGACCACGGCCAGGGCCTGTGGCCGAACCTGCGGGTGGCCGA 375

Qy        579 GCTCACCCTGGACGACTACATCTACCCCTGGAGGAACCTCGCCCTAGACATGGCCAACGG 638
              || || || | ||| | |   || ||    |.|||  | ||  | |  |  |||     |
Db        376 GCGCATCCAGCACGTCGAACGCTTCCTGGCGCGGATGCGCGAGCAACGCGAGGCGGTGGT 435

Qy        639 CGTGCT 644
              |  |||
Db        436 CAAGCT 441


Search completed: March 18, 2004, 05:51:21
Job time : 62.9049 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 02:39:53 ; Search time 235.943 Seconds
(without alignments)
16085.819 Million cell updates/sec

Title: US-09-664-186-4
Perfect score: 1026
Sequence: 1 gtgaagaacgaaaaaacctt.....ccctcaaggaggcctcctga 1026

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2438257 seqs, 1849576744 residues

Total number of hits satisfying chosen parameters: 4876514

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	53.8	5.2	9025608	14	US-10-156-761-1	Sequence 1, Appli
	2	53.2	5.2	1755	15	US-10-260-238-721	Sequence 721, App
	3	51.6	5.0	1815	9	US-09-758-269-13	Sequence 13, Appl
	4	50.8	5.0	1018	12	US-10-425-114-35941	Sequence 35941, A
	5	50.8	5.0	1077	14	US-10-156-761-6573	Sequence 6573, Ap
	6	50.8	5.0	9025608	14	US-10-156-761-1	Sequence 1, Appli
	7	50	4.9	675	14	US-10-156-761-6076	Sequence 6076, Ap
	8	49.8	4.9	1389	14	US-10-156-761-5431	Sequence 5431, Ap
	9	49.6	4.8	1674	12	US-10-282-122A-26289	Sequence 26289, A
	10	49.6	4.8	1677	12	US-10-282-122A-28540	Sequence 28540, A
	11	49.4	4.8	898	12	US-10-425-114-32815	Sequence 32815, A
	12	49.2	4.8	1803	15	US-10-369-493-42225	Sequence 42225, A
	13	48.8	4.8	1045	12	US-10-425-114-26418	Sequence 26418, A
	14	48.6	4.7	960	15	US-10-369-493-34210	Sequence 34210, A
	15	48.6	4.7	9369	14	US-10-200-562-190	Sequence 190, App
	16	48.6	4.7	9369	14	US-10-237-551-190	Sequence 190, App
	17	48.6	4.7	9369	14	US-10-237-551-247	Sequence 247, App
c	18	48.6	4.7	154746	10	US-09-827-688-8	Sequence 8, Appli
	19	48	4.7	1977	14	US-10-355-430-17	Sequence 17, Appl
	20	47.8	4.7	963	12	US-10-282-122A-11900	Sequence 11900, A
	21	47.6	4.6	669	12	US-10-425-114-34873	Sequence 34873, A
	22	47.6	4.6	864	12	US-10-425-114-35445	Sequence 35445, A
	23	47.6	4.6	1527	14	US-10-156-761-6366	Sequence 6366, Ap
	24	47.4	4.6	1926	14	US-10-156-761-2557	Sequence 2557, Ap
	25	47.2	4.6	969	14	US-10-156-761-111	Sequence 111, App
	26	46.8	4.6	1224	14	US-10-156-761-4891	Sequence 4891, Ap
	27	46.8	4.6	1605	15	US-10-260-238-908	Sequence 908, App
	28	46.8	4.6	2223	14	US-10-156-761-3605	Sequence 3605, Ap
	29	46.6	4.5	1071	15	US-10-369-493-42984	Sequence 42984, A
	30	46.6	4.5	10065	14	US-10-160-758-1	Sequence 1, Appli
	31	46.6	4.5	10065	15	US-10-085-117-114	Sequence 114, App
	32	46.6	4.5	10455	14	US-10-160-758-4	Sequence 4, Appli
	33	46.6	4.5	11073	14	US-10-160-758-2	Sequence 2, Appli
	34	46.6	4.5	11073	15	US-10-085-117-113	Sequence 113, App
	35	46.4	4.5	1381	12	US-10-425-114-13514	Sequence 13514, A
	36	46.2	4.5	1414	12	US-10-425-114-14100	Sequence 14100, A
c	37	46.2	4.5	1473	9	US-09-853-161-43	Sequence 43, Appl
c	38	46.2	4.5	1473	9	US-09-852-659A-43	Sequence 43, Appl
c	39	46.2	4.5	1473	9	US-09-852-797-43	Sequence 43, Appl
	40	46.2	4.5	2892	14	US-10-156-761-2448	Sequence 2448, Ap
	41	46	4.5	808	12	US-10-424-599-52912	Sequence 52912, A
	42	46	4.5	1263	15	US-10-369-493-40502	Sequence 40502, A
	43	46	4.5	1418	12	US-10-425-114-1960	Sequence 1960, Ap
	44	46	4.5	1532	12	US-10-425-114-34918	Sequence 34918, A
	45	46	4.5	1864	12	US-10-425-114-28403	Sequence 28403, A

ALIGNMENTS

RESULT 1
US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match 5.2%; Score 53.8; DB 14; Length 9025608;
Best Local Similarity 45.0%; Pred. No. 0.00014;
Matches 202; Conservative 0; Mismatches 247; Indels 0; Gaps 0;

```
Qy       261 ACCCCCCGTTGGGAATGTATGTCAAGCCGTCCAGGACGGGCCTCAGAAGCTTCTGGAACT 320
             |||| |   | | | |  | ||  ||||||  |  | |  ||  | |||  |  |  ||
Db   4922513 ACCCGTCCCCGCGCAGGGTTATCGTGCCGTCATGACCCGCTCTGTGGAGCGGCCCGTCCT 4922454
```

```
Qy        321 CCTCCAGGAGATTGCCCGCTCCACCATCCCCTACGGCAACCGGGAGCTCTGGAGGAAGGT 380
Db    4922453 TTTCCTCGACGTCGACGGACCCCTCATCCCGTTCGGGCCGTCGTCCCGGCGGCTCCAGCC 4922394

Qy        381 GGGGACGGTCGTCTTCATGGTCCCCCTGGAGATGTTGGCCCTCAACCTGGGGGTCACCCG 440
Db    4922393 CCGTGCCGCCGGCTCCGTGGCGTTCTCCGACGAGGGGAACCCGCTGCTGGGGCGGCTCGA 4922334

Qy        441 GCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGAAAAAGGGCCTGGTGGCCACCGACGT 500
Db    4922333 CCAGGCCGTCGGACCGCGTCTGATGGCACTGGGGTGTGAGCTGGTGTGGGCCTCGACCTG 4922274

Qy        501 CCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCATCGGCACCCTTTGGGCCGTCCGGCT 560
Db    4922273 GATGAATGACGCCAACGAGTCCGTCGCCCCACGGATCGGACTGCCCAGGCTGCCTGTGGT 4922214

Qy        561 GAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACTACATCTACCCCTGGAGGAACCTCGC 620
Db    4922213 GGAGTGGCCGGAAGCCTCCACCCACGAGGGCCCGCGCGGCCTGCACTGGAAGACCCGCCA 4922154

Qy        621 CCTAGACATGGCCAACGGCGTGCTCTCCTTCAACTGGGTCAAGGCCTACCAGGACCACGG 680
Db    4922153 TCTCGTCGAGTGGGCCGGCCACCGCCGTTCATCTGGGTCGACGACGAGATCGGCGCCAT 4922094

Qy        681 AATCCGCCCCACCCTGGACGTGCTGGTCC 709
Db    4922093 GGACCGCCTCTGGGTCGCCGCGCAGCACC 4922065
```

RESULT 2
US-10-260-238-721
; Sequence 721, Application US/10260238
; Publication No. US20040016025A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, Paul R.
; APPLICANT: Moughamer, Todd G.
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiyaki
; APPLICANT: Kreps, Joel
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
; FILE REFERENCE: 60111-NP
; CURRENT APPLICATION NUMBER: US/10/260,238
; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/325,448
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 6077
; SEQ ID NO 721
; LENGTH: 1755
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: N_region
; LOCATION: (917)..(917)
; OTHER INFORMATION: n = any nucleotide
; FEATURE:
; NAME/KEY: N_region
; LOCATION: (1750)..(1750)
; OTHER INFORMATION: n = any nucleotide
US-10-260-238-721

Query Match 5.2%; Score 53.2; DB 15; Length 1755;
Best Local Similarity 50.8%; Pred. No. 1.9e-05;
Matches 127; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

```
Qy        543 CCTTTGGGCCGTCCGGCTGAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACTACATCTA 602
Db        246 CGTCTTCGATGGCCGGGAGCTCGCCCGGAACGACAGCCGCATCATGGACTACGGCGTCTC 305

Qy        603 CCCCTGGAGGAACCTCGCCCTAGACATGGCCAACGGCGTGCTCTCCTTCAACTGGGTCAA 662
Db        306 CCACGGCAACGTCCTCCACCTCGTCATCCGCATCTCCGACCTCCGCCTCATCACCGTCCA 365

Qy        663 GGCCTACCAGGACCACGGAATCCGCCCCACCCTGGACGTGCTGGTCCTCTGGGCTCAGGG 722
Db        366 GACCGTCCACGGCAACAAGTTCAGGTTCCGCGTCGAGCCCGGCCGCACCGTCGGCTACGT 425

Qy        723 GAAAAGGGTGATGCCCAACACCAAGACCGTGGCCGTTGACCTGGGCCTCATCCTGGTCCT 782
Db        426 CAAGCAGCAGATCGCCAAGAACTCCACCCACGACGACGACCACCACAGCCTCGTCCTCCA 485

Qy        783 CCCCGAGGTG 792
Db        486 GGGCGAGGTG 495
```

RESULT 3
US-09-758-269-13
; Sequence 13, Application US/09758269
; Patent No. US20020104120A1
; GENERAL INFORMATION:
; APPLICANT: IUCHI, SATOSHI
; APPLICANT: KOBAYASHI, MASATOMO
; APPLICANT: SHINOZAKI, KAZUO
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN
; TITLE OF INVENTION: CLEAVAGE ENZYME GENE
; FILE REFERENCE: 3914-3
; CURRENT APPLICATION NUMBER: US/09/758,269
; CURRENT FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: JP 2001-003476
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: JP 2000-010056
; PRIOR FILING DATE: 2000-01-13
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 1815
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1812)
US-09-758-269-13

Query Match 5.0%; Score 51.6; DB 9; Length 1815;
Best Local Similarity 45.9%; Pred. No. 5.8e-05;
Matches 177; Conservative 0; Mismatches 209; Indels 0; Gaps 0;

```
Qy        320 TCCTCCAGGAGATTGCCCGCTCCACCATCCCCTACGGCAACCGGGAGCTCTGGAGGAAGG 379
Db        155 TCGCCGACCTGCCTGCGCCGTCCAGGAAGCCCGCCGCCATTGCCGTCCCAGGGCACGCCG 214

Qy        380 TGGGGACGGTCGTCTTCATGGTCCCCCTGGAGATGTTGGCCCTCAACCTGGGGGTCACCC 439
Db        215 CGGCGCCGAGGAAAGCGGAGGGCGGCAAGAAGCAGCTCAACTTGTTCCAGCGCGCCGCGG 274

Qy        440 GGCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGAAAAAGGGCCTGGTGGCCACCGACG 499
Db        275 CGGCCGCGCTCGACGCGTTCGAGGAAGGGTTCGTGGCCAACGTCCTCGAGCGGCCCCACG 334

Qy        500 TCCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCATCGGCACCCTTTGGGCCGTCCGGC 559
Db        335 GGCTGCCCAGCACGGCCGACCCGGCCGTGCAGATCGCCGGCAACTTCGCGCCCGTCGGGG 394

Qy        560 TGAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACTACATCTACCCCTGGAGGAACCTCG 619
```

```
Db        395 AGAGGCCGCCCGTGCACGAGCTCCCCGTCTCCGGCCGCATCCCGCCCTTCATCGACGGGG 454

Qy        620 CCCTAGACATGGCCAACGGCGTGCTCTCCTTCAACTGGGTCAAGGCCTACCAGGACCACG 679

Db        455 TCTACGCGCGCAACGGCGCCAACCCCTGCTTCGACCCCGTCGCGGGGCACCACCTCTTCG 514

Qy        680 GAATCCGCCCCACCCTGGACGTGCTG 705

Db        515 ACGGCGACGGCATGGTGCACGCGCTG 540


RESULT 4
US-10-425-114-35941
; Sequence 35941, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 35941
;   LENGTH: 1018
;   TYPE: DNA
;   ORGANISM: Zea mays subsp. mexicana
;   FEATURE:
;   OTHER INFORMATION: Clone ID: UC-ZMROTEOSINTE020D09_FLI
US-10-425-114-35941

  Query Match             5.0%;  Score 50.8;  DB 12;  Length 1018;
  Best Local Similarity   46.6%;  Pred. No. 8.6e-05;
  Matches  163;  Conservative    0;  Mismatches  187;  Indels    0;  Gaps    0;

Qy        284 AAGCCGTCCAGGACGGGCCTCAGAAGCTTCTGGAACTCCTCCAGGAGATTGCCCGCTCCA 343

Db        389 AGGCCGCCAAGCACAAGGCCGGCGAGACGACGGAGGCCGCCAAGCAGAAGGCCGGCGAGA 448

Qy        344 CCATCCCCTACGGCAACCGGGAGCTCTGGAGGAAGGTGGGGACGGTCGTCTTCATGGTCC 403

Db        449 CGACGGAGGCCGCCAAGCAGAAGGCCGGCGAGACGACGGAGGCCGCCAAGCAGAAGGCAG 508

Qy        404 CCCTGGAGATGTTGGCCCTCAACCTGGGGGTCACCCGGCAGACCGTCCACGCCTGGAAGA 463

Db        509 CCGACGCCATGGAGGCCGCCAAGCAGAAGGCCGCCGAGGCCGGGCAGTACGCCAAGGACA 568

Qy        464 AGGTCCTTGAGAAAAAGGGCCTGGTGGCCACCGACGTCCTTCACCAAACCGTCAACGGGG 523

Db        569 CCGCCGTCTCCGGCAAGGACAAGTCCGGCGGCGTCATCCAGCAGGCCACTGAGCAGGTGA 628

Qy        524 AGCGCCGGGCCATCGGCACCCTTTGGGCCGTCCGGCTGAGGCCAGGGAAAGCCAGGCTCA 583

Db        629 AGAGCGCGGCGGCGGGCGCCAAGGACGCGGTGATGAGCACGCTGGGGATGGGCGGGGACA 688

Qy        584 CCCTGGACGACTACATCTACCCCTGGAGGAACCTCGCCCTAGACATGGCC 633

Db        689 ACAAGCAGGGCGACGCCAACACCAACACTAACACCAACACCAACAAGGAC 738


RESULT 5
US-10-156-761-6573
; Sequence 6573, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
;  APPLICANT:  IKEDA, HARUO
;  APPLICANT:  ISHIKAWA, JUN
;  APPLICANT:  HORIKAWA, HIROSHI
;  APPLICANT:  SHIBA, TADAYOSHI
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  APPLICANT:  HATTORI, MASAHIRA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-262
;  CURRENT APPLICATION NUMBER: US/10/156,761
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 6573
;   LENGTH: 1077
;   TYPE: DNA
;   ORGANISM: Streptomyces avermitilis
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1077)
US-10-156-761-6573

  Query Match             5.0%;  Score 50.8;  DB 14;  Length 1077;
  Best Local Similarity   44.3%;  Pred. No. 8.7e-05;
  Matches  208;  Conservative    0;  Mismatches  262;  Indels    0;  Gaps    0;

Qy        323 TCCAGGAGATTGCCCGCTCCACCATCCCCTACGGCAACCGGGAGCTCTGGAGGAAGGTGG 382

Db        560 TCGGCGGGGTCGGCCTCGCGACGCTCCAGGCCGCCCGGATCGCGGGCGCGTCGAAGATCG 619

Qy        383 GGACGGTCGTCTTCATGGTCCCCCTGGAGATGTTGGCCCTCAACCTGGGGGTCACCCGGC 442

Db        620 TCGCGGTGGACGTCTCGCCGGAGAAGGAGTCGCTCGCCCGCCGGGCCGGGGCCACCGACT 679

Qy        443 AGACCGTCCACGCCTGGAAGAAGGTCCTTGAGAAAAAGGGCCTGGTGGCCACCGACGTCC 502

Db        680 ACCTCGTCGCCTCCGAGAACACGGCCCGCGAGATCCGCGGGCTCACCGACAAGCAGGGGG 739

Qy        503 TTCACCAAACCGTCAACGGGGAGCGCCGGGCCATCGGCACCCTTTGGGCCGTCCGGCTGA 562

Db        740 TGGACGTCGCCGTCGAGTGCGTGGGCCGCGCCGTGACCATCCGCACCGCCTGGGACTCGA 799

Qy        563 GGCCAGGGAAAGCCAGGCTCACCCTGGACGACTACATCTACCCCTGGAGGAACCTCGCCC 622

Db        800 CGCGGCGTGGCGGCCGCACCACGGTCGTCGGCATCGGCGGCAAGGACCAGCAGGTCGTGT 859

Qy        623 TAGACATGGCCAACGGCGTGCTCTCCTTCAACTGGGTCAAGGCCTACCAGGACCACGGAA 682

Db        860 TCAACGCGCTGGAGATCTTCCACTGGGGCCGCACCCTGTCGGGCTGCGTGTACGGCAACT 919

Qy        683 TCCGCCCCACCCTGGACGTGCTGGTCCTCTGGGCTCAGGGGAAAAGGGTGATGCCCAACA 742

Db        920 CGGACCCGGCGAAGGACCTGCCGGTCCTGGCCGAGCACGTACGGGCGGGGCGCCTGGACC 979

Qy        743 CCAAGACCGTGGCCGTTGACCTGGGCCTCATCCTGGTCCTCCCCGAGGTG 792

Db        980 TCGGCGCGCTGGTGACGGAGCGGATCGCCCTCGACGGCATCCCGGCGGCG 1029


RESULT 6
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
;  APPLICANT:  IKEDA, HARUO
;  APPLICANT:  ISHIKAWA, JUN
;  APPLICANT:  HORIKAWA, HIROSHI
;  APPLICANT:  SHIBA, TADAYOSHI
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  APPLICANT:  HATTORI, MASAHIRA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-262
```

```
;  CURRENT APPLICATION NUMBER: US/10/156,761
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
;   LENGTH: 9025608
;   TYPE: DNA
;   ORGANISM: Streptomyces avermitilis
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4187715)
;   OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

  Query Match             5.0%;  Score 50.8;  DB 14;  Length 9025608;
  Best Local Similarity   44.3%;  Pred. No. 0.0011;
  Matches  208;  Conservative    0;  Mismatches  262;  Indels    0;  Gaps    0;

Qy        323 TCCAGGAGATTGCCCGCTCCACCATCCCCTACGGCAACCGGGAGCTCTGGAGGAAGGTGG 382
Db    7895416 TCGGCGGGGTCGGCCTCGCGACGCTCCAGGCCGCCCGGATCGCGGGCGCGTCGAAGATCG 7895475

Qy        383 GGACGGTCGTCTTCATGGTCCCCCTGGAGATGTTGGCCCTCAACCTGGGGGTCACCCGGC 442
Db    7895476 TCGCGGTGGACGTCTCGCCGGAGAAGGAGTCGCTCGCCCGCCGGGCCGGGGCCACCGACT 7895535

Qy        443 AGACCGTCCACGCCTGGAAGAAGGTCCTTGAGAAAAAGGGCCTGGTGGCCACCGACGTCC 502
Db    7895536 ACCTCGTCGCCTCCGAGAACACGGCCCGCGAGATCCGCGGGCTCACCGACAAGCAGGGGG 7895595

Qy        503 TTCACCAAACCGTCAACGGGGAGCGCCGGGCCATCGGCACCCTTTGGGCCGTCCGGCTGA 562
Db    7895596 TGGACGTCGCCGTCGAGTGCGTGGGCCGCGCCGTGACCATCCGCACCGCCTGGGACTCGA 7895655

Qy        563 GGCCAGGGAAAGCCAGGCTCACCCTGGACGACTACATCTACCCCTGGAGGAACCTCGCCC 622
Db    7895656 CGCGGCGTGGCGGCCGCACCACGGTCGTCGGCATCGGCGGCAAGGACCAGCAGGTCGTGT 7895715

Qy        623 TAGACATGGCCAACGGCGTGCTCTCCTTCAACTGGGTCAAGGCCTACCAGGACCACGGAA 682
Db    7895716 TCAACGCGCTGGAGATCTTCCACTGGGGCCGCACCCTGTCGGGCTGCGTGTACGGCAACT 7895775

Qy        683 TCCGCCCCACCCTGGACGTGCTGGTCCTCTGGGCTCAGGGGAAAAGGGTGATGCCCAACA 742
Db    7895776 CGGACCCGGCGAAGGACCTGCCGGTCCTGGCCGAGCACGTACGGGCGGGGCGCCTGGACC 7895835

Qy        743 CCAAGACCGTGGCCGTTGACCTGGGCCTCATCCTGGTCCTCCCCGAGGTG 792
Db    7895836 TCGGCGCGCTGGTGACGGAGCGGATCGCCCTCGACGGCATCCCGGCGGCG 7895885


RESULT 7
US-10-156-761-6076
; Sequence 6076, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
;  APPLICANT:  IKEDA, HARUO
;  APPLICANT:  ISHIKAWA, JUN
;  APPLICANT:  HORIKAWA, HIROSHI
;  APPLICANT:  SHIBA, TADAYOSHI
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  APPLICANT:  HATTORI, MASAHIRA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-262
;  CURRENT APPLICATION NUMBER: US/10/156,761
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 6076
;   LENGTH: 675
;   TYPE: DNA
;   ORGANISM: Streptomyces avermitilis
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(675)
US-10-156-761-6076

  Query Match             4.9%;  Score 50;  DB 14;  Length 675;
  Best Local Similarity   44.7%;  Pred. No. 0.00013;
  Matches  237;  Conservative    0;  Mismatches  290;  Indels    3;  Gaps    1;

Qy        388 GTCGTCTTCATGGTCCCCCTGGAGATGTTGGCCCTCAACCTGGGGGTCACCCGGCAGACC 447
Db        106 GACCTCGCCGAGTTCCTCGACGCGGCCCTGGCCGGCGGCGTCGACATCGTGCAGCTGCGC 165

Qy        448 GTCCACGCCTGGAAGAAGGTCCTTGAGAAAAAGGGCCTGGTGGCCACCGACGTCCTTCAC 507
Db        166 GACAAGGGCATGGAGGCGGCCGAGGAGCTGGAGCACCTCCAGGTCTTCGCGGAC---GCC 222

Qy        508 CAAACCGTCAACGGGGAGCGCCGGGCCATCGGCACCCTTTGGGCCGTCCGGCTGAGGCCA 567
Db        223 TGTCGCCGCCACGGCAAGCTCCTCGCGGTCAACGACCGGGCGGACGTCGCGCACGCCGTC 282

Qy        568 GGGAAAGCCAGGCTCACCCTGGACGACTACATCTACCCCTGGAGGAACCTCGCCCTAGAC 627
Db        283 GACTCCGACGTGCTCCACCTGGGGCAGGGCGACCTCCCCGTCCCGGCGGCCCGGGCGATC 342

Qy        628 ATGGCCAACGGCGTGCTCTCCTTCAACTGGGTCAAGGCCTACCAGGACCACGGAATCCGC 687
Db        343 CTCGGCGCCGACGTCCTCATCGGCCGCTCCACGCATGCCGAGGCCGAGGCCGAGGCCGCC 402

Qy        688 CCCACCCTGGACGTGCTGGTCCTCTGGGCTCAGGGGAAAAGGGTGATGCCCAACACCAAG 747
Db        403 GCCGTCCAGGAGGGCGTGGACTACTTCTGTACGGGTCCCTGCTGGCCCACCCCCACCAAG 462

Qy        748 ACCGTGGCCGTTGACCTGGGCCTCATCCTGGTCCTCCCCGAGGTGGAGCGTTCCAAACTC 807
Db        463 CCCGGCCGGCACGCCCCGGGCCTCGACCTGGTCCGCCACACGGCCGCCCTGGGCACCGAC 522

Qy        808 CCGGCCCTTATCACCCTCATTGCTACGTACATTGCCGATCTCCTAGATGACCGTCGTTCA 867
Db        523 CGCCCCTGGTTCGCCATCGGCGGCATCGACCTCGGCAATCTCGACGAGGTTCTCGAAGCG 582

Qy        868 AGACGTTTCTATGCAGGCTTGCTGTGGGCTGTGGCCAGGGGTGAACTCCC 917
Db        583 GGCGCCCGCCGTGTCGTCGTCGTGCGCGCGATCACCGAGGCGGACGACCC 632


RESULT 8
US-10-156-761-5431
; Sequence 5431, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
;  APPLICANT:  IKEDA, HARUO
;  APPLICANT:  ISHIKAWA, JUN
;  APPLICANT:  HORIKAWA, HIROSHI
;  APPLICANT:  SHIBA, TADAYOSHI
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  APPLICANT:  HATTORI, MASAHIRA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-262
;  CURRENT APPLICATION NUMBER: US/10/156,761
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
```

; SEQ ID NO 5431
; LENGTH: 1389
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1389)
US-10-156-761-5431

Query Match 4.9%; Score 49.8; DB 14; Length 1389;
Best Local Similarity 46.2%; Pred. No. 0.00019;
Matches 165; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

```
Qy        438 CCGGCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGAAAAAGGGCCTGGTGGCCACCGA 497
Db        957 CGGCCAGATGGGCCTCGACCTCGAGAAGTACCTCGAGATCCAGGGCAAGACGGTCGAGGA 1016

Qy        498 CGTCCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCATCGGCACCCTTTGGGCCGTCCG 557
Db       1017 GTTCGACGCCGAGACCAAGGAAGCGGCGGTCAAGGGCATCAAGACGCAGTTCGTCCTCGA 1076

Qy        558 GCTGAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACTACATCTACCCCTGGAGGAACCT 617
Db       1077 CGAGCTCGTCAACAAGGAGAAGCTCAACGTCAACCAGGAGGAGCTCACCGAGCACCTCAT 1136

Qy        618 CGCCCTAGACATGGCCAACGGCGTGCTCTCCTTCAACTGGGTCAAGGCCTACCAGGACCA 677
Db       1137 GCGGCGTGCGGCTTCCTCCGGCATGTCCCCCGACCAGTTCGCCCAGGCGGTCGTCGAGGG 1196

Qy        678 CGGAATCCGCCCCACCCTGGACGTGCTGGTCCTCTGGGCTCAGGGGAAAAGGGTGATGCC 737
Db       1197 TGGCCAGGTCCCGATGCTGGTCGGCGAGGTCGCCCGCGGCAAGGCGCTGGCCGTCGTGGT 1256

Qy        738 CAACACCAAGACCGTGGCCGTTGACCTGGGCCTCATCCTGGTCCTCCCCGAGGTGGA 794
Db       1257 CGAGGCCGCCACGGTGAAGGACACCAACGGCGAGATCGTCGACCTGGACGACGAGGA 1313
```

RESULT 9
US-10-282-122A-26289
; Sequence 26289, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 26289
; LENGTH: 1674
; TYPE: DNA
; ORGANISM: Mycobacterium bovis
US-10-282-122A-26289

Query Match 4.8%; Score 49.6; DB 12; Length 1674;
Best Local Similarity 45.7%; Pred. No. 0.00023;
Matches 212; Conservative 0; Mismatches 249; Indels 3; Gaps 1;

```
Qy        391 GTCTTCATGGTCCCCCTGGAGATGTTGGCCCTCAACCTGGGGGTCACCCGGCAGACCGTC 450
Db        184 GCCTTCCTGGCCACCGGCGCCACCGTGGGCATCCTGCAACAGGAACCGCCGCTGAACGAG 243

Qy        451 CACGCCTGGAAGAAGGTCCTTGAGAAAAAGGGCCTGGTGGCCACCGACGTCCTTCACCAA 510
Db        244 GACAAGACCGTTCGCGGCAATGTGGAAGAGGGCATGGGGGACATCAAGATCAAGCTCGAC 303

Qy        511 ACCGTCAACGGGGAGCGCCGGGCCATCGGCACCCTTTGGGCCGTCCGGCTGAGGCCAGGG 570
Db        304 CGCTTCAACGAGGTCGCCGAATTGATGGCCACCGACTACACCGACGAGCTGATGGAAGAG 363

Qy        571 AAAGCCAGGCTCACCCTGGACGACTACATCTACCCCTGGAGGAACCTCGCCCTAGACATG 630
Db        364 ATGGGTCGGCTGCAAGAGGAACTGGACCACGCCGACGCGTGGGACCTCGACGCGCAGCTC 423

Qy        631 GCCAACGGCGTGCTCTCCTTCAACTGGGTCAAGGCCTACCAGGACCACGGAATCCGCCCC 690
Db        424 GAGCAGGCCATGGATGCGCTGCGCTGTCCGCCGGCCGAC---GAGCCGGTAACCAACCTA 480

Qy        691 ACCCTGGACGTGCTGGTCCTCTGGGCTCAGGGGAAAAGGGTGATGCCCAACACCAAGACC 750
Db        481 TCCGGTGGCGAGCGTCGCCGGGTGGCGCTGTGCAAACTGCTGTTGTCCAAACCCGACCTG 540

Qy        751 GTGGCCGTTGACCTGGGCCTCATCCTGGTCCTCCCCGAGGTGGAGCGTTCCAAACTCCCG 810
Db        541 TTGTTGCTCGACGAGCCGACCAACCACCTAGACGCGGAAAGTGTGCAGTGGCTCGAACAG 600

Qy        811 GCCCTTATCACCCTCATTGCTACGTACATTGCCGATCTCCTAGA 854
Db        601 CATCTGGCCAGCTACCCCGGTGCGATCCTGGCGGTCACCCACGA 644
```

RESULT 10
US-10-282-122A-28540
; Sequence 28540, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 28540
; LENGTH: 1677
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-10-282-122A-28540

Query Match 4.8%; Score 49.6; DB 12; Length 1677;
Best Local Similarity 45.7%; Pred. No. 0.00023;
Matches 212; Conservative 0; Mismatches 249; Indels 3; Gaps 1;

```
Qy        391 GTCTTCATGGTCCCCCTGGAGATGTTGGCCCTCAACCTGGGGGTCACCCGGCAGACCGTC 450
Db        184 GCCTTCCTGGCCACCGGCGCCACCGTGGGCATCCTGCAACAGGAACCGCCGCTGAACGAG 243

Qy        451 CACGCCTGGAAGAAGGTCCTTGAGAAAAAGGGCCTGGTGGCCACCGACGTCCTTCACCAA 510
Db        244 GACAAGACCGTTCGCGGCAATGTGGAAGAGGGCATGGGGGACATCAAGATCAAGCTCGAC 303

Qy        511 ACCGTCAACGGGGAGCGCCGGGCCATCGGCACCCTTTGGGCCGTCCGGCTGAGGCCAGGG 570
Db        304 CGCTTCAACGAGGTCGCCGAATTGATGGCCACCGACTACACCGACGAGCTGATGGAAGAG 363

Qy        571 AAAGCCAGGCTCACCCTGGACGACTACATCTACCCCTGGAGGAACCTCGCCCTAGACATG 630
Db        364 ATGGGTCGGCTGCAAGAGGAACTGGACCACGCCGACGCGTGGGACCTCGACGCGCAGCTC 423

Qy        631 GCCAACGGCGTGCTCTCCTTCAACTGGGTCAAGGCCTACCAGGACCACGGAATCCGCCCC 690
Db        424 GAGCAGGCCATGGATGCGCTGCGCTGTCCGCCGGCCGAC---GAGCCGGTAACCAACCTA 480

Qy        691 ACCCTGGACGTGCTGGTCCTCTGGGCTCAGGGGAAAAGGGTGATGCCCAACACCAAGACC 750
Db        481 TCCGGTGGCGAGCGTCGCCGGGTGGCGCTGTGCAAACTGCTGTTGTCCAAACCCGACCTG 540

Qy        751 GTGGCCGTTGACCTGGGCCTCATCCTGGTCCTCCCCGAGGTGGAGCGTTCCAAACTCCCG 810
Db        541 TTGTTGCTCGACGAGCCGACCAACCACCTAGACGCGGAAAGTGTGCAGTGGCTCGAACAG 600

Qy        811 GCCCTTATCACCCTCATTGCTACGTACATTGCCGATCTCCTAGA 854
Db        601 CATCTGGCCAGCTACCCCGGTGCGATCCTGGCGGTCACCCACGA 644
```

RESULT 11
US-10-425-114-32815
; Sequence 32815, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 32815
; LENGTH: 898
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLMO17024H02_FLI
US-10-425-114-32815

Query Match 4.8%; Score 49.4; DB 12; Length 898;
Best Local Similarity 44.0%; Pred. No. 0.00022;
Matches 209; Conservative 0; Mismatches 266; Indels 0; Gaps 0;

```
Qy        330 GATTGCCCGCTCCACCATCCCCTACGGCAACCGGGAGCTCTGGAGGAAGGTGGGGACGGT 389
Db          1 GATCGAGGACTGCTACATCGTCTCCGGCGACGACTGCGTGGCCGTGAAGAGCGGGTGGGA 60

Qy        390 CGTCTTCATGGTCCCCCTGGAGATGTTGGCCCTCAACCTGGGGGTCACCCGGCAGACCGT 449
Db         61 CGAGTACGGCATCAGGTTCAACATGCCGAGCCAGCACATCGTCATCCGGAGGCTGACCTG 120

Qy        450 CCACGCCTGGAAGAAGGTCCTTGAGAAAAAGGGCCTGGTGGCCACCGACGTCCTTCACCA 509
Db        121 CGTCTCCCCCACGAGCGCCATGATCGCGCTGGGCAGCGAGATGTCCGGCGGCATCCGCGA 180

Qy        510 AACCGTCAACGGGGAGCGCCGGGCCATCGGCACCCTTTGGGCCGTCCGGCTGAGGCCAGG 569
Db        181 CGTGCGCGCCGAGGACAGCGTCGCCATCAACACGGAGTCGGCCGTCAGGGTCAAGTCCGG 240

Qy        570 GAAAGCCAGGCTCACCCTGGACGACTACATCTACCCCTGGAGGAACCTCGCCCTAGACAT 629
Db        241 TGCCGGGAGGGGCGGCTTCGTCAAGGACATCTTCGTGCGCGGCCTCAGCCTCCACACCAT 300

Qy        630 GGCCAACGGCGTGCTCTCCTTCAACTGGGTCAAGGCCTACCAGGACCACGGAATCCGCCC 689
Db        301 GAAGTGGGTGTTCTGGATGACCGGCAACTACGGGCAGCACCCCGACAACACGTCCGACCC 360

Qy        690 CACCCTGGACGTGCTGGTCCTCTGGGCTCAGGGGAAAAGGGTGATGCCCAACACCAAGAC 749
Db        361 CAACGCCATGCCCGAGGTCACCGGCATCAACTACAGCGACGTGTTCGCCGAGAACGTGAC 420

Qy        750 CGTGGCCGTTGACCTGGGCCTCATCCTGGTCCTCCCCGAGGTGGAGCGTTCCAAA 804
Db        421 CATGGCCGGCAGGATGGAGGGCATCCCTAAGGACCCCTACACCGGGATCTGCATA 475
```

RESULT 12
US-10-369-493-42225
; Sequence 42225, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374

```
; SEQ ID NO 42225
;   LENGTH: 1803
;   TYPE: DNA
;   ORGANISM: Halobacterium sp. NRC-1
US-10-369-493-42225

  Query Match             4.8%;  Score 49.2;  DB 15;  Length 1803;
  Best Local Similarity   48.3%;  Pred. No. 0.0003;
  Matches  138;  Conservative    0;  Mismatches  148;  Indels    0;  Gaps    0;

Qy        426 CCTGGGGGTCACCCGGCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAGAAAAAGGGCCT 485
Db        198 CCTCCTCGTCACCGACTTCATCGACCGCGCGCGCACCTACTACGGCGACGAAACCGCCGT 257

Qy        486 GGTGGCCACCGACGTCCTTCACCAAACCGTCAACGGGGAGCGCCGGGCCATCGGCACCCT 545
Db        258 CGTCGCCACCACCGGCCGGCGGTACACGTACAGCGAGCTCGCCGACCGCGCCGACCGCTT 317

Qy        546 TTGGGCCGTCCGGCTGAGGCCAGGGAAAGCCAGGCTCACCCTGGACGACTACATCTACCC 605
Db        318 TTCGGCGGTCCTCCAGGCGGCCGGCATCGACAGCGGCGACCGCGTCGCCGTCCTCGACCC 377

Qy        606 CTGGAGGAACCTCGCCCTAGACATGGCCAACGGCGTGCTCTCCTTCAACTGGGTCAAGGC 665
Db        378 GAACACCCACTACCACCTCGAAGCCGCCTACGGCGCGATGCAGCTGGGGGCCATCCACAC 437

Qy        666 CTACCAGGACCACGGAATCCGCCCCACCCTGGACGTGCTGGTCCTC 711
Db        438 GCCACTGAACTACCGGCTCACGCCCGACGACTACGCGTACATCCTC 483

RESULT 13
US-10-425-114-26418
; Sequence 26418, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 26418
;   LENGTH: 1045
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: LIB4172-028-B4_FLI
US-10-425-114-26418

  Query Match             4.8%;  Score 48.8;  DB 12;  Length 1045;
  Best Local Similarity   45.7%;  Pred. No. 0.00034;
  Matches  170;  Conservative    0;  Mismatches  202;  Indels    0;  Gaps    0;

Qy        334 GCCCGCTCCACCATCCCCTACGGCAACCGGGAGCTCTGGAGGAAGGTGGGGACGGTCGTC 393
Db        384 GCGCCGTCCAGGAAGCCCGCCGCCATTGCCGTCCCAGGGCACTCCGCGGCGCCGAGGAAA 443

Qy        394 TTCATGGTCCCCCTGGAGATGTTGGCCCTCAACCTGGGGGTCACCCGGCAGACCGTCCAC 453
Db        444 GCGGACGGCGGCAAGAAGCAGCTCAACTTGTTCCAGCGCGCCGCGGCGGCCGCGCTCGAC 503

Qy        454 GCCTGGAAGAAGGTCCTTGAGAAAAAGGGCCTGGTGGCCACCGACGTCCTTCACCAAACC 513
Db        504 GCGTTCGAGGAAGGGTTCGTGGCCAACGTCCTGGAGCGGCCCCACGGGCTGCCCAGCACG 563

Qy        514 GTCAACGGGGAGCGCCGGGCCATCGGCACCCTTTGGGCCGTCCGGCTGAGGCCAGGGAAA 573
Db        564 GCCGACCCGGCCGTGCAGATCGCCGGCAACTTCGCGCCCGTCGGGGAGAGGCCGCCCGTG 623

Qy        574 GCCAGGCTCACCCTGGACGACTACATCTACCCCTGGAGGAACCTCGCCCTAGACATGGCC 633
Db        624 CACGAGCTCCCCGTCTCCGGCCGCATCCCGCCCTTCATCGACGGGGTCTACGCGCGCAAC 683

Qy        634 AACGGCGTGCTCTCCTTCAACTGGGTCAAGGCCTACCAGGACCACGGAATCCGCCCCACC 693
Db        684 GGCGCCAACCCCTGCTTCGACCCCGTCGCCGGGCACCACCTCTTCGACGGCGACGGCATG 743

Qy        694 CTGGACGTGCTG 705
Db        744 GTGCACGCGCTG 755

RESULT 14
US-10-369-493-34210
; Sequence 34210, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
;  APPLICANT: Cao, Yongwei
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Chen, Xianfeng
;  TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;  TITLE OF INVENTION:  PLANTS WITH IMPROVED PROPERTIES
;  FILE REFERENCE: 38-10(52052)B
;  CURRENT APPLICATION NUMBER: US/10/369,493
;  CURRENT FILING DATE:  2003-02-28
;  PRIOR APPLICATION NUMBER: US 60/360,039
;  PRIOR FILING DATE: 2002-02-21
;  NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 34210
;   LENGTH: 960
;   TYPE: DNA
;   ORGANISM: Sphingomonas aromaticivorans
US-10-369-493-34210

  Query Match             4.7%;  Score 48.6;  DB 15;  Length 960;
  Best Local Similarity   45.0%;  Pred. No. 0.00038;
  Matches  183;  Conservative    0;  Mismatches  224;  Indels    0;  Gaps    0;

Qy        396 CATGGTCCCCCTGGAGATGTTGGCCCTCAACCTGGGGGTCACCCGGCAGACCGTCCACGC 455
Db        204 CATCGCAGGCGCGGACGTGATCATCGTCACCGCCGGCGTCGCCCGCAAGCCCGGCATGAG 263

Qy        456 CTGGAAGAAGGTCCTTGAGAAAAAGGGCCTGGTGGCCACCGACGTCCTTCACCAAACCGT 515
Db        264 CCGCGACGACCTGCTCGGCATCAACCTCAAGGTGATGAAGGCCGTCGGCGAAGGCATCCG 323

Qy        516 CAACGGGGAGCGCCGGGCCATCGGCACCCTTTGGGCCGTCCGGCTGAGGCCAGGGAAAGC 575
Db        324 CGACAACGCGCCCGACGCGTTCGTGATCTGCATCACCAACCCGCTCGACGCGATGGTCTG 383

Qy        576 CAGGCTCACCCTGGACGACTACATCTACCCCTGGAGGAACCTCGCCCTAGACATGGCCAA 635
Db        384 GGCGCTGCGCGAGTTCTCGGGCCTGCCCGCGAACAAGGTCGTCGGCATGGCCGGCGTGCT 443

Qy        636 CGGCGTGCTCTCCTTCAACTGGGTCAAGGCCTACCAGGACCACGGAATCCGCCCCACCCT 695
Db        444 CGACTCGGCGCGCTTCAGCACGTTCCTGGCATGGGAATTCGGCGTCTCGATCCGCGACGT 503

Qy        696 GGACGTGCTGGTCCTCTGGGCTCAGGGGAAAAGGGTGATGCCCAACACCAAGACCGTGGC 755
Db        504 GAACACGTTCGTTCTCGGCGGCCACGGCGACACCATGGTTCCGGTCACCCAGTACTCGAC 563

Qy        756 CGTTGACCTGGGCCTCATCCTGGTCCTCCCCGAGGTGGAGCGTTCCA 802
Db        564 CGTCAACGGCATCCCGGTGCCCGACCTCGTCAAGATGGGCCTGTCCA 610
```

RESULT 15
US-10-200-562-190
; Sequence 190, Application US/10200562
; Publication No. US20030165819A1
; GENERAL INFORMATION:
; APPLICANT: McGowen, Patrick
; APPLICANT: Hosken, Nancy A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
; TITLE OF INVENTION: TREATMENT OF HERPES SIMPLEX VIRUS INFECTION
; FILE REFERENCE: 210121.538C2
; CURRENT APPLICATION NUMBER: US/10/200,562
; CURRENT FILING DATE: 2002-07-19
; NUMBER OF SEQ ID NOS: 212
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 190
; LENGTH: 9369
; TYPE: DNA
; ORGANISM: HSV2
US-10-200-562-190

Query Match 4.7%; Score 48.6; DB 14; Length 9369;
Best Local Similarity 43.9%; Pred. No. 0.00073;
Matches 207; Conservative 0; Mismatches 264; Indels 0; Gaps 0;

```
Qy        355 GGCAACCGGGAGCTCTGGAGGAAGGTGGGGACGGTCGTCTTCATGGTCCCCCTGGAGATG 414
              | ||| |   ||||||  ||     || |  | ||| || |  | ||||   || || |
Db       3643 GCCATCGACGAGCTCACGACCCTGGGGTCGGCGGCCGCCGACCTGGTAGTCCAGGCGCGC 3702

Qy        415 TTGGCCCTCAACCTGGGGGTCACCCGGCAGACCGTCCACGCCTGGAAGAAGGTCCTTGAG 474
                ||||  |  || |  ||        ||  ||| || | |       || ||  | |
Db       3703 CGGGCCTGCGCCCAGCCGGAGGGGGACCATGCCGCCCTCATCGATGCCGCGGCCCGCGCG 3762

Qy        475 AAAAAGGGCCTGGTGGCCACCGACGTCCTTCACCAAACCGTCAACGGGGAGCGCCGGGCC 534
              |  |  |     ||  | |  ||     ||| |  | |    |||||  |  | |  |
Db       3763 ACGACCGCTGCTCGGGAAAGCCTCGCGGGGCACGAGGCTGGGTTCGGGGGCCTGCTGCAC 3822

Qy        535 ATCGGCACCCTTTGGGCCGTCCGGCTGAGGCCAGGGAAAGCCAGGCTCACCCTGGACGAC 594
               ||           |  | ||           |||   |||  ||      |||| | |
Db       3823 GCCGAAGGGACGGCCGGGGACCATTCCCCCAGCGGGCGCGCCCTGCAGGAGCTGGGCAAG 3882

Qy        595 TACATCTACCCCTGGAGGAACCTCGCCCTAGACATGGCCAACGGCGTGCTCTCCTTCAAC 654
                ||||  | ||  | |   || ||||   ||  | |     || |||   |  | | |
Db       3883 GTCATCGGCGCCACGCGACGCCGCGCCGACGAACTTGAGGCCGCCGTCGCCGACCTGACG 3942

Qy        655 TGGGTCAAGGCCTACCAGGACCACGGAATCCGCCCCACCCTGGACGTGCTGGTCCTCTGG 714
              |   | |||   ||||  |  | |   | ||  |   |       |     | | ||  |
Db       3943 GCGAAGATGGCGGCCCAGCGCGCCCGGGGCAGCAGCGAGCGCTGGGCCGCCGGCGTCGAG 4002

Qy        715 GCTCAGGGGAAAAGGGTGATGCCCAACACCAAGACCGTGGCCGTTGACCTGGGCCTCATC 774
              ||  |  | |  | ||  |  |  | || ||  ||  |  || || ||| ||| | |
Db       4003 GCGGCGCTGGACCGCGTCGAGAACCGCGCCGAGTTCGACGTGGTCGAGCTGCGCCGCCTG 4062

Qy        775 CTGGTCCTCCCCGAGGTGGAGCGTTCCAAACTCCCGGCCCTTATCACCCTC 825
              | ||  ||  | |    | |  | | ||| | || || | |   ||  | |
Db       4063 CAGGCGCTGGCGGGCACGCACGGCTACAACCCCCGGGACTTCCGCAAGCGC 4113
```

Search completed: March 18, 2004, 10:21:33
Job time : 259.943 secs